
Journal of Medical Internet Research

Journal Impact Factor (JIF) (2023): 5.8
Volume 24 (2022), Issue 3 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

The Effectiveness of Mobile Phone Messaging–Based Interventions to Promote Physical Activity in Type 2 Diabetes Mellitus: Systematic Review and Meta-analysis (e29663) Mohammed Alsahli, Alaa Abd-Alrazaq, Mowafa Househ, Stathis Konstantinidis, Holly Blake.	8
Standards, Processes, and Tools Used to Evaluate the Quality of Health Information Systems: Systematic Literature Review (e26577) René Noël, Carla Taramasco, Gastón Márquez.	20
Community Health Programs Delivered Through Information and Communications Technology in High-Income Countries: Scoping Review (e26515) Hannah Beks, Olivia King, Renee Clapham, Laura Alston, Kristen Glenister, Carol McKinstry, Claire Quilliam, Ian Wellwood, Catherine Williams, Anna Wong Shee.	31
Implementation of Cognitive Behavioral Therapy in e–Mental Health Apps: Literature Review (e27791) Kerstin Denecke, Nicole Schmid, Stephan Nüssli.	49
Improving the Development and Implementation of Audit and Feedback Systems to Support Health Care Workers in Limiting Antimicrobial Resistance in the Hospital: Scoping Review (e33531) Julia Keizer, Britt Bente, Nashwan Al Naiemi, Lisette Van Gemert-Pijnen, Nienke Beerlage-De Jong.	62
Use of Mobile and Wearable Artificial Intelligence in Child and Adolescent Psychiatry: Scoping Review (e33560) Victoria Welch, Tom Wy, Anna Ligezka, Leslie Hassett, Paul Croarkin, Arjun Athreya, Magdalena Romanowicz.	85
Workarounds in Electronic Health Record Systems and the Revised Sociotechnical Electronic Health Record Workaround Analysis Framework: Scoping Review (e33046) Vincent Blijleven, Florian Hoxha, Monique Jaspers.	97
Effects of a Nurse-Led Telehealth Self-care Promotion Program on the Quality of Life of Community-Dwelling Older Adults: Systematic Review and Meta-analysis (e31912) Arkers Wong, Jonathan Bayuo, Frances Wong, Wing Yuen, Athena Lee, Pui Chang, Jojo Lai.	113
Inequities in Health Care Services Caused by the Adoption of Digital Health Technologies: Scoping Review (e34144) Rui Yao, Wenli Zhang, Richard Evans, Guang Cao, Tianqi Rui, Lining Shen.	130
Success Factors of Medical Crowdfunding Campaigns: Systematic Review (e30189) Xiaoyan Hou, Tailai Wu, Zhuo Chen, Liqin Zhou.	146

Process and Outcome Evaluations of Smartphone Apps for Bipolar Disorder: Scoping Review ([e29114](#))
 Iona Tatham, Ellisiv Clarke, Kelly Grieve, Pulkit Kaushal, Jan Smeddinck, Evelyn Millar, Aditya Sharma. 158

Digital Assessment Tools Using Animation Features to Quantify Alcohol Consumption: Systematic App Store and Literature Review ([e28927](#))
 Veronika Wiemker, Maria Neufeld, Anna Bunova, Ina Danquah, Carina Ferreira-Borges, Stefan Konigorski, Ankit Rastogi, Charlotte Probst. 1 7 0

The Effectiveness of Physical Activity-Promoting Web- and Mobile-Based Distance Weight Loss Interventions on Body Composition in Rehabilitation Settings: Systematic Review, Meta-analysis, and Meta-Regression Analysis ([e25906](#))
 Heli Lahtio, Aki Rintala, Jaakko Immonen, Tuulikki Sjögren. 188

Role of Telemedicine in Inflammatory Bowel Disease: Systematic Review and Meta-analysis of Randomized Controlled Trials ([e28978](#))
 Lanlan Pang, Hengyu Liu, Zhidong Liu, Jinyu Tan, Long-yuan Zhou, Yun Qiu, Xiaoqing Lin, Jinshen He, Xuehua Li, Sinan Lin, Subrata Ghosh, Ren Mao, Minhu Chen. 207

Value Cocreation in Health Care: Systematic Review ([e33061](#))
 Yuxin Peng, Tailai Wu, Zhuo Chen, Zhaohua Deng. 222

Telemedicine Acceptance Among Older Adult Patients With Cancer: Scoping Review ([e28724](#))
 Ning-Qi Pang, Jerrald Lau, Si-Ying Fong, Celine Wong, Ker-Kan Tan. 234

Telehealth Business Models and Their Components: Systematic Review ([e33128](#))
 Farnia Velayati, Haleh Ayatollahi, Morteza Hemmat, Reza Dehghan. 250

Synthesizing Dimensions of Digital Maturity in Hospitals: Systematic Review ([e32994](#))
 Rhona Duncan, Rebekah Eden, Leanna Woods, Ides Wong, Clair Sullivan. 266

Physician Burnout and the Electronic Health Record Leading Up to and During the First Year of COVID-19: Systematic Review ([e36200](#))
 Clemens Kruse, Michael Mileski, Gevin Dray, Zakia Johnson, Cameron Shaw, Harsha Shirodkar. 277

Digital Education for Health Professionals: An Evidence Map, Conceptual Framework, and Research Agenda ([e31977](#))
 Lorainne Tudor Car, Selina Poon, Bhone Kyaw, David Cook, Victoria Ward, Rifat Atun, Azeem Majeed, Jamie Johnston, Rianne van der Kleij, Mariam Molokhia, Florian V Wangenheim, Martin Lupton, Niels Chavannes, Onyema Ajuebor, Charles Prober, Josip Car. 860

Viewpoints

Patient Influencers: The Next Frontier in Direct-to-Consumer Pharmaceutical Marketing ([e29422](#))
 Erin Willis, Marjorie Delbaere. 294

How Can Research on Artificial Empathy Be Enhanced by Applying Deepfakes? ([e29506](#))
 Hsuan-Chia Yang, Annisa Rahmanti, Chih-Wei Huang, Yu-Chuan Li. 303

Tackling the Burden of Electronic Health Record Use Among Physicians in a Mental Health Setting: Physician Engagement Strategy ([e32800](#))
 Tania Tajirian, Damian Jankowicz, Brian Lo, Lydia Sequeira, Gillian Strudwick, Khaled Almilaji, Vicky Stergiopoulos. 311

A National Network of Safe Havens: Scottish Perspective (e31684)	
Chuang Gao, Mark McGilchrist, Shahzad Mumtaz, Christopher Hall, Lesley Anderson, John Zurowski, Sharon Gordon, Joanne Lumsden, Vicky Munro, Artur Wozniak, Michael Sibley, Christopher Banks, Chris Duncan, Pamela Linksted, Alastair Hume, Catherine Stables, Charlie Mayor, Jacqueline Caldwell, Katie Wilde, Christian Cole, Emily Jefferson.	322
Scale-up of Digital Innovations in Health Care: Expert Commentary on Enablers and Barriers (e24582)	
Hannes Schlieter, Lisa Marsch, Diane Whitehouse, Lena Otto, Ana Londral, Gisbert Teepe, Martin Benedict, Joseph Ollier, Tom Ulmer, Nathalie Gasser, Sabine Ultsch, Bastian Wollschlaeger, Tobias Kowatsch.	338
The Disclosure of Personally Identifiable Information in Studies of Neighborhood Contexts and Patient Outcomes (e30619)	
Andrew Rundle, Michael Bader, Stephen Mooney.	349
Gray Literature in Evaluating Effectiveness in Digital Health and Health and Welfare Technology: A Source Worth Considering (e29307)	
Sara Landerdahl Stridsberg, Matt Richardson, Ken Redekop, Maria Ehn, Sarah Wamala Andersson.	354
 Original Papers	
Increasing the Effectiveness of a Physical Activity Smartphone Intervention With Positive Suggestions: Randomized Controlled Trial (e32130)	
Aleksandrina Skvortsova, Talia Cohen Rodrigues, David de Buissonjé, Tobias Kowatsch, Prabhakaran Santhanam, Dieuwke Veldhuijzen, Henriët van Middendorp, Andrea Evers.	366
Effect of Source Type and Protective Message on the Critical Evaluation of News Messages on Facebook: Randomized Controlled Trial in the Netherlands (e27945)	
Frans Folkvord, Freek Snelling, Doeschka Anschutz, Tilo Hartmann, Alexandra Theben, Laura Gunderson, Ivar Vermeulen, Francisco Lupiáñez-Villanueva.	381
Cost-effectiveness of Web-Based and Home-Based Postnatal Psychoeducational Interventions for First-time Mothers: Economic Evaluation Alongside Randomized Controlled Trial (e25821)	
Qishi Zheng, Luming Shi, Lixia Zhu, Nana Jiao, Yap Chong, Sally Chan, Yiong Chan, Nan Luo, Wenru Wang, Honggu He.	392
Clinical Outcomes Among Working Adults Using the Health Integrator Smartphone App: Analyses of Prespecified Secondary Outcomes in a Randomized Controlled Trial (e24725)	
Stephanie Bonn, Gabriella Licitra, Rino Bellocco, Ylva Trolle Lagerros.	408
Social Media–Delivered Patient Education to Enhance Self-management and Attitudes of Patients with Type 2 Diabetes During the COVID-19 Pandemic: Randomized Controlled Trial (e31449)	
Cheng Leong, Ting-I Lee, Yu-Mei Chien, Li-Na Kuo, Yu-Feng Kuo, Hsiang-Yin Chen.	419
Review of Mobile Apps for Women With Anxiety in Pregnancy: Maternity Care Professionals' Guide to Locating and Assessing Anxiety Apps (e31831)	
Kerry Evans, Jasper Donelan, Stefan Rennick-Egglestone, Serena Cox, Yvonne Kuipers.	433
Efficacy, Benefits, and Harms of a Self-management App in a Swedish Trauma-Exposed Community Sample (PTSD Coach): Randomized Controlled Trial (e31419)	
Ida Hensler, Josefin Sveen, Martin Cernvall, Filip Arnberg.	444
Effect of a Lifestyle-Focused Web-Based Application on Risk Factor Management in Patients Who Have Had a Myocardial Infarction: Randomized Controlled Trial (e25224)	
Halldóra Ögmundsdóttir Michelsen, Ingela Sjölin, Maria Bäck, Manuel Gonzalez Garcia, Anneli Olsson, Camilla Sandberg, Alexandru Schiopu, Margrét Leósdóttir.	459

The Characteristics and Functionalities of Mobile Apps Aimed at Patients Diagnosed With Immune-Mediated Inflammatory Diseases: Systematic App Search (e31016)	
Rosa Romero-Jimenez, Vicente Escudero-Vilaplana, Esther Chamorro-De-Vega, Arantza Ais-Larisgoitia, Maria Lobato Matilla, Ana Herranz-Alonso, Maria Sanjurjo.	477
An mHealth Intervention to Reduce the Packing of Discretionary Foods in Children’s Lunch Boxes in Early Childhood Education and Care Services: Cluster Randomized Controlled Trial (e27760)	
Nicole Pearson, Meghan Finch, Rachel Sutherland, Melanie Kingsland, Luke Wolfenden, Taya Wedesweiler, Vanessa Herrmann, Sze Yoong.	4 8 7
Education on Depression in Mental Health Apps: Systematic Assessment of Characteristics and Adherence to Evidence-Based Guidelines (e28942)	
Laura Martinengo, Anne-Claire Stona, Lorainne Tudor Car, Jimmy Lee, Konstadina Griva, Josip Car.	501
Effect of Sleep Disturbance Symptoms on Treatment Outcome in Blended Cognitive Behavioral Therapy for Depression (E-COMPARED Study): Secondary Analysis (e30231)	
Esben Jensen, Nicolai Ladegaard, Angelina Mellentin, David Ebert, Ingrid Titzler, Ricardo Araya, Arlinda Cerga Pashoja, Jean-Baptiste Hazo, Jérôme Holtzmann, Roman Cieslak, Ewelina Smoktunowicz, Rosa Baños, Rocio Herrero, Azucena García-Palacios, Cristina Botella, Thomas Berger, Tobias Krieger, Trine Holmberg, Naira Topooco, Gerhard Andersson, Annemieke van Straten, Lise Kemmeren, Annet Kleiboer, Heleen Riper, Kim Mathiasen.	517
Internet-Delivered Cognitive Behavioral Therapy for Generalized Anxiety Disorder in Nationwide Routine Care: Effectiveness Study (e29384)	
Ville Ritola, Jari Lipsanen, Satu Pihlaja, Eero-Matti Gummerus, Jan-Henry Stenberg, Suoma Saarni, Grigori Joffe.	530
The Perceived Benefits of Digital Interventions for Behavioral Health: Qualitative Interview Study (e34300)	
Gabriela Marcu, Steven Ondersma, Allison Spiller, Brianna Broderick, Reema Kadri, Lorraine Buis.	546
Barriers and Considerations in the Design and Implementation of Digital Behavioral Interventions: Qualitative Analysis (e34301)	
Gabriela Marcu, Steven Ondersma, Allison Spiller, Brianna Broderick, Reema Kadri, Lorraine Buis.	559
Social-Cyber Maneuvers During the COVID-19 Vaccine Initial Rollout: Content Analysis of Tweets (e34040)	
Janice Blane, Daniele Bellutta, Kathleen Carley.	576
Understanding Public Perceptions of Per- and Polyfluoroalkyl Substances: Infodemiology Study of Social Media (e25614)	
Hao Tian, Christy Gaines, Lori Launi, Ana Pomales, Germaine Vazquez, Amanda Goharian, Bradley Goodnight, Erica Haney, Christopher Reh, Rachel Rogers.	591
The Public Perception of the #GeneEditedBabies Event Across Multiple Social Media Platforms: Observational Study (e31687)	
Congning Ni, Zhiyu Wan, Chao Yan, Yongtai Liu, Ellen Clayton, Bradley Malin, Zhijun Yin.	605
Puff Bars, Tobacco Policy Evasion, and Nicotine Dependence: Content Analysis of Tweets (e27894)	
Kar-Hai Chu, Tina Hershey, Beth Hoffman, Riley Wolynn, Jason Colditz, Jaime Sidani, Brian Primack.	621
Content Analysis of Nicotine Poisoning (Nic Sick) Videos on TikTok: Retrospective Observational Infodemiology Study (e34050)	
Vidya Purushothaman, Tiana McMann, Matthew Nali, Zhuoran Li, Raphael Cuomo, Tim Mackey.	631
Exploring the Discursive Emphasis on Patients and Coaches Who Participated in Technology-Assisted Diabetes Self-management Education: Clinical Implementation Study of Health360x (e23535)	
Muhammed Idris, Ernest Alema-Mensah, Elizabeth Olorundare, Mohammad Mohammad, Michelle Brown, Elizabeth Ofili, Priscilla Pemu.	6 3 9

Modeling Access Across the Digital Divide for Intersectional Groups Seeking Web-Based Health Information: National Survey (e32678)
 Kristina Medero, Kelly Merrill Jr, Morgan Ross. 659

Enabling Eating Detection in a Free-living Environment: Integrative Engineering and Machine Learning Study (e27934)
 Bo Zhang, Kaiwen Deng, Jie Shen, Lingrui Cai, Bohdana Ratitch, Haoda Fu, Yuanfang Guan. 680

Tracking Subjective Sleep Quality and Mood With Mobile Sensing: Multiverse Study (e25643)
 Koen Niemeijer, Merijn Mestdag, Peter Kuppens. 696

Demographic and Psychosocial Characteristics Associated With Use of a Prostate Cancer Survivorship Website: Implications From a Multisite Randomized Controlled Trial (e27890)
 Allison Marziliano, Michael Diefenbach, Shawna Hudson, Erin Tagai, Elizabeth Handorf, Alicja Bator, Suzanne Miller. 713

Validity Evidence of the eHealth Literacy Questionnaire (eHLQ) Part 2: Mixed Methods Approach to Evaluate Test Content, Response Process, and Internal Structure in the Australian Community Health Setting (e32777)
 Christina Cheng, Gerald Elsworth, Richard Osborne. 724

Comparing SF-36 Scores Collected Through Web-Based Questionnaire Self-completions and Telephone Interviews: An Ancillary Study of the SENTIPAT Multicenter Randomized Controlled Trial (e29009)
 Ay e Açma, Fabrice Carrat, Gilles Hejblum. 741

Development of a Prognostic App (iCanPredict) to Predict Survival for Chinese Women With Breast Cancer: Retrospective Study (e35768)
 Zhuo Ma, Sijia Huang, Xiaoqing Wu, Yinying Huang, Sally Chan, Yilan Lin, Xujuan Zheng, Jiemin Zhu. 755

Machine Learning Prediction Models for Gestational Diabetes Mellitus: Meta-analysis (e26634)
 Zheqing Zhang, Luqian Yang, Wentao Han, Yaoyu Wu, Linhui Zhang, Chun Gao, Kui Jiang, Yun Liu, Huiqun Wu. 768

The Mediating Role of Patients' Trust Between Web-Based Health Information Seeking and Patients' Uncertainty in China: Cross-sectional Web-Based Survey (e25275)
 Wei Dong, Xiangxi Lei, Yongmei Liu. 783

Social Media Images as an Emerging Tool to Monitor Adherence to COVID-19 Public Health Guidelines: Content Analysis (e24787)
 Sean Young, Qingpeng Zhang, Daniel Zeng, Yongcheng Zhan, William Cumberland. 806

An mHealth Intervention to Improve Medication Adherence and Health Outcomes Among Patients With Coronary Heart Disease: Randomized Controlled Trial (e27202)
 Zhao Ni, Bei Wu, Qing Yang, Lijing Yan, Changqing Liu, Ryan Shaw. 811

An Innovative Telemedical Network to Improve Infectious Disease Management in Critically Ill Patients and Outpatients (TELnet@NRW): Stepped-Wedge Cluster Randomized Controlled Trial (e34098)
 Gernot Marx, Wolfgang Greiner, Christian Juhra, Svenja Elkenkamp, Daniel Gensorowsky, Sebastian Lemmen, Jan Englbrecht, Sandra Dohmen, Antje Gottschalk, Miriam Haverkamp, Annette Hempen, Christian Flügel-Bleienheuft, Daniela Bause, Henna Schulze-Steinen, Susanne Rademacher, Jennifer Kistermann, Stefan Hoch, Hans-Juergen Beckmann, Christian Lanckohr, Volker Lowitsch, Arne Peine, Fabian Juzek-Kuepper, Carina Benstoem, Kathrin Sperling, Robert Deisz. 827

Effectiveness, Cost-effectiveness, and Cost-Utility of a Digital Smoking Cessation Intervention for Cancer Survivors: Health Economic Evaluation and Outcomes of a Pragmatic Randomized Controlled Trial (e27588)
 Ajla Mujcic, Matthijs Blankers, Brigitte Boon, Irma Verdonck-de Leeuw, Filip Smit, Margriet van Laar, Rutger Engels. 843

A Question-and-Answer System to Extract Data From Free-Text Oncological Pathology Reports (CancerBERT Network): Development Study ([e27210](#))
Joseph Mitchell, Phillip Szepietowski, Rachel Howard, Phillip Reisman, Jennie Jones, Patricia Lewis, Brooke Fridley, Dana Rollison. 881

Privacy Preservation in Patient Information Exchange Systems Based on Blockchain: System Design Study ([e29108](#))
Sejong Lee, Jaehyeon Kim, Yongseok Kwon, Teasung Kim, Sunghyun Cho. 898

Improving Diagnosis Through Digital Pathology: Proof-of-Concept Implementation Using Smart Contracts and Decentralized File Storage ([e34207](#))
Hemang Subramanian, Susmitha Subramanian. 918

A Web-Based Self-assessment Model for Evaluating Multidisciplinary Cancer Teams in Spain: Development and Validation Pilot Study ([e29063](#))
Mercedes Guilbert, Joan Prades, Josep Borrás, Inmaculada Maestu, Juan Guerra, Lluís Fumadó, José Mira, AEMAC Program Research Team. 935

Contextualizing Engagement With Health Information on Facebook: Using the Social Media Content and Context Elicitation Method ([e25243](#))
Yonaira Rivera, Meghan Moran, Johannes Thrul, Corinne Joshu, Katherine Smith. 950

Influence of Forced Online Distance Education During the COVID-19 Pandemic on the Perceived Stress of Postsecondary Students: Cross-sectional Study ([e30778](#))
Andrej Šorgo, Nuša Crnkovi, Branko Gabrovec, Katarina Cesar, Špela Selak. 987

The Impact of COVID-19 Lockdown on Daily Activities, Cognitions, and Stress in a Lonely and Distressed Population: Temporal Dynamic Network Analysis ([e32598](#))
Matthias Haucke, Andreas Heinz, Shuyan Liu, Stephan Heinzl. 1000

Stratified Impacts of the Infodemic During the COVID-19 Pandemic: Cross-sectional Survey in 6 Asian Jurisdictions ([e31088](#))
Xi Chen, Fen Lin, Edmund Cheng. 1012

The Evaluation of a Social Media Campaign to Increase COVID-19 Testing in Migrant Groups: Cluster Randomized Trial ([e34544](#))
Ingeborg Elgersma, Atle Fretheim, Thor Indseth, Anita Munch, Live Johannessen, Christine Hansen. 1032

Tracking Public Attitudes Toward COVID-19 Vaccination on Tweets in Canada: Using Aspect-Based Sentiment Analysis ([e35016](#))
Hyeju Jang, Emily Rempel, Ian Roe, Prince Adu, Giuseppe Carenini, Naveed Janjua. 1042

The Challenge of Debunking Health Misinformation in Dynamic Social Media Conversations: Online Randomized Study of Public Masking During COVID-19 ([e34831](#))
Mehdi Mourali, Carly Drake. 1052

Letters to the Editor

Toward a Better Understanding of Quality Social Connections. Comment on “Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis” ([e36739](#))
Huachu Deng, Xingan Qin. 962

The Unclear Role of the Physician on Social Media During the COVID-19 Pandemic. Comment on “Emergency Physician Twitter Use in the COVID-19 Pandemic as a Potential Predictor of Impending Surge: Retrospective Observational Study” (e34870)
 Spencer MacLeod, Nikhi Singh, Carter Boyd. 964

Interpretation Bias Toward the Positive Impacts of Digital Interventions in Health Care. Comment on “Value of the Electronic Medical Record for Hospital Care: Update From the Literature” (e37208)
 Erfan Shakibaei Bonakdeh. 966

Authors’ Reply to: Interpretation Bias Toward the Positive Impacts of Digital Interventions in Health Care. Comment on “Value of the Electronic Medical Record for Hospital Care: Update From the Literature” (e37419)
 Jürgen Stausberg, Aykut Uslu. 968

Authors’ Reply to: Toward a Better Understanding of Quality Social Connections. Comment on “Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis” (e37440)
 Lindsay Dewa, Lily Roberts, Emma Lawrance, Hutun Ashrafian. 971

Challenges in Measuring What Matters to Patients With Diabetes. Comment on “Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review” (e36876)
 Femke Rutters, Ellen Elsmann, Lenka Groeneveld, Marlous Langendoen-Gort, Lidwine Mokkink, Caroline Terwee. 973

Authors’ Reply to: Challenges in Measuring What Matters to Patients With Diabetes. Comment on “Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review” (e37957)
 Yu Kwan, Jie Phang, Sungwon Yoon, Lian Low. 975

Corrigenda and Addendas

Correction: Using Participatory Design Methodologies to Co-Design and Culturally Adapt the Spanish Version of the Mental Health eClinic: Qualitative Study (e37679)
 Laura Ospina-Pinillos, Tracey Davenport, Antonio Mendoza Diaz, Alvaro Navarro-Mancilla, Elizabeth Scott, Ian Hickie. 978

Correction: COVID-19 Vaccine Tweets After Vaccine Rollout: Sentiment-Based Topic Modeling (e37841)
 Luwen Huangfu, Yiwen Mo, Peijie Zhang, Daniel Zeng, Saikhe He. 980

Metadata Correction: Clinical Outcomes Among Working Adults Using the Health Integrator Smartphone App: Analyses of Prespecified Secondary Outcomes in a Randomized Controlled Trial (e38199)
 Stephanie Bonn, Gabriella Licitra, Rino Bellocco, Ylva Trolle Lagerros. 983

Correction: Gender-Specific Impact of Self-Monitoring and Social Norm Information on Walking Behavior Among Chinese College Students Assessed Using WeChat: Longitudinal Tracking Study (e38221)
 Yuepei Xu, Ling-Zi Yue, Wei Wang, Xiao-Ju Wu, Zhu-Yuan Liang. 985

Review

The Effectiveness of Mobile Phone Messaging–Based Interventions to Promote Physical Activity in Type 2 Diabetes Mellitus: Systematic Review and Meta-analysis

Mohammed Alsahli^{1,2}, PhD; Alaa Abd-Alrazaq^{3,4}, PhD; Mowafa Househ³, PhD; Stathis Konstantinidis^{1,5}, PhD; Holly Blake^{1,5}, PhD

¹School of Health Sciences, Faculty of Medicine and Health Sciences, University of Nottingham, Nottingham, United Kingdom

²Division of Health Informatics, College of Health Science, Saudi Electronic University, Riyadh, Saudi Arabia

³Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Qatar Foundation, Doha, Qatar

⁴AI Center for Precision Health, Weill Cornell Medicine-Qatar, Doha, Qatar

⁵Nottingham Biomedical Research Centre, National Institute for Health Research, Nottingham, United Kingdom

Corresponding Author:

Mowafa Househ, PhD

Division of Information and Computing Technology, College of Science and Engineering

Hamad Bin Khalifa University

Qatar Foundation

PO Box 34110, Doha Al Luqta St

Ar-Rayyan

Doha, 11023/2093

Qatar

Phone: 974 55708549

Email: mhouseh@hbku.edu.qa

Abstract

Background: The prevalence of type 2 diabetes mellitus (T2DM) is increasing worldwide. Physical activity (PA) is an important aspect of self-care and first line management for T2DM. SMS text messaging can be used to support self-management in people with T2DM, but the effectiveness of mobile text message–based interventions in increasing PA is still unclear.

Objective: This study aims to assess the effectiveness of mobile phone messaging on PA in people with T2DM by summarizing and pooling the findings of previous literature.

Methods: A systematic review was conducted to accomplish this objective. Search sources included 5 bibliographic databases (MEDLINE, Cochrane Library, CINAHL, Web of Science, and Embase), the search engine *Google Scholar* (Google Inc), and backward and forward reference list checking of the included studies and relevant reviews. A total of 2 reviewers (MA and AA) independently carried out the study selection, data extraction, risk of bias assessment, and quality of evidence evaluation. The results of the included studies were synthesized narratively and statistically, as appropriate.

Results: We included 3.8% (6/151) of the retrieved studies. The results of individual studies were contradictory regarding the effectiveness of mobile text messaging on PA. However, a meta-analysis of the results of 5 studies showed no statistically significant effect ($P=.16$) of text messages on PA in comparison with no intervention. A meta-analysis of the findings of 2 studies showed a nonsignificant effect ($P=.14$) of text messages on glycemic control. Of the 541 studies, 2 (0.4%) found a nonsignificant effect of text messages on anthropometric measures (weight and BMI).

Conclusions: We could not draw a definitive conclusion regarding the effectiveness of text messaging on PA, glycemic control, weight, or BMI among patients with T2DM, given the limited number of included studies and their high risk of bias. Therefore, there is a need for more high-quality primary studies.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020156465; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=156465

(*J Med Internet Res* 2022;24(3):e29663) doi:[10.2196/29663](https://doi.org/10.2196/29663)

KEYWORDS

type 2 diabetes mellitus; physical activity; mobile phone messaging; systematic review; meta-analysis

Introduction

Background

The burden of diabetes is increasing, and the number of people with type 2 diabetes mellitus (T2DM) worldwide has reached 387 million and is expected to increase to 592 million by 2035 [1]. This prevalence imposes a high and rising burden of lifelong multiorgan complications, leading to increased disability and risk of premature deaths, mainly in low- and middle-income countries [2]. A considerable amount of literature suggests that better management of T2DM delays the onset of short- and long-term complications among people diagnosed with T2DM [3-5]. Over the past decades, physical activity (PA) has been part of the first line T2DM care management [6]. PA includes all movements that increase energy use; however, there are three main types of exercise: aerobic, strength training, and flexibility work [7]. PA can help people with T2DM achieve a variety of goals, including increased vigor, improved glycemic hemoglobin control, decreased insulin resistance, increased cardiorespiratory fitness, improved lipid profile, blood pressure reduction, and maintenance of weight loss [8]. Unfortunately, patients with T2DM are less likely to engage in regular PA, with recent estimates demonstrating a lower participation rate compared with the national average [9]. There have been many attempts to explore alternative approaches to improve PA in people with T2DM, and the mobile phone messaging revolution has brought entirely new opportunities and increased access to self-management education [1]. The literature shows that text messaging-based interventions can be effective in improving health-related behaviors and bridging the gaps between patients and health care services for people living with chronic diseases [10,11]. Text messaging may be 1-way (unidirectional) or 2-way (bidirectional); they can be standardized or tailored to specific patients and sent at varied frequencies based on the intervention design [12]. Multiple meta-analyses have demonstrated the overall success of mobile phone messaging in promoting various aspects of behavior change for PA and mental health-related disorders [1,13,14].

Research Problem and Aim

Several studies have assessed the effect of mobile text messaging on the PA of patients with T2DM. It is crucial to summarize and aggregate the findings of such studies to produce more generalizable and definitive conclusions about the effectiveness of such interventions. A total of 4 previous systematic reviews did not provide evidence from studies with text messaging interventions that specifically targeted PA. Specifically, the first review focused on the impact of education on T2DM delivered via mobile text messaging [15]. The second review assessed the effectiveness of text messaging interventions on glycated hemoglobin (HbA_{1c}) in patients with T2DM, including all self-management strategies [1]. The third review identified randomized trials conducted to improve glycemic control in T2DM, which involved the delivery of behavior change content through a range of digital platforms and approaches (eg, SMS

text messaging, multimedia message service, or instant messaging such as WhatsApp) [12]. The fourth review assessed the effectiveness of technology-based interventions to promote PA in T2DM; for this review, technology included mobile phones and text messages, websites, CD-ROMs, and computer learning-based technology [16]. This review was conducted approximately 7 years ago, but studies involving technology-based interventions are rapidly emerging and there may be new published evidence. Therefore, this study aims to assess the effectiveness of mobile phone messaging on PA in patients with T2DM by summarizing and pooling the findings of previous literature.

Methods

Overview

A systematic review was conducted and reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (Multimedia Appendix 1) [17]. The protocol for this review was registered at PROSPERO (ID: CRD42020156465).

Search Strategy

Search Sources

We used the following electronic databases in our search: MEDLINE, Cochrane Library, CINAHL, Web of Science, and Embase. These databases were searched on April 19, 2020, by the first author (MA). Auto alerts were set after searching the databases to conduct an automatic search weekly for 16 weeks (ending on August 9, 2020) and send us the retrieved studies. We also searched the search engine *Google Scholar* (Google Inc) to identify gray literature. To identify further studies of relevance to the review, we screened the reference lists of included studies (ie, backward reference list checking) and identified and screened studies that cited the included studies (ie, forward reference list checking).

Search Terms

The search terms were identified by consulting 2 experts in eHealth interventions for patients with diabetes and by checking systematic reviews of relevance to the review. These terms were chosen based on the target population (eg, type 2 diabetes, diabetes type 2, and type II diabetes), target intervention (eg, text messaging, text messages, and short messages), target outcome (eg, PA, physical exercise, HbA_{1c}, and weight), and target study design (eg, trial, experiment, and randomized controlled trial [RCT]). Multimedia Appendix 2 shows the detailed search query used for searching MEDLINE.

Study Eligibility Criteria

The population of interest was adult patients (≥18 years) with T2DM, regardless of their gender and ethnicity. We excluded patients with type 1 diabetes mellitus, gestational diabetes, and prediabetes. The target intervention in this review was mobile phone text messages (SMS text messaging and multimedia

message service), but not mobile apps, web-delivered interventions, wearables, or emails. The aim of the text messages was to improve solely PA but not diet, lifestyle, diabetic literacy, or other aspects of self-care. The primary outcomes of interest were subjectively or objectively measured PA (eg, step counts), glycemic control (eg, HbA_{1c} and fasting glucose), and anthropometric measures (eg, change in weight and BMI). Only RCTs were eligible for inclusion in this review. We considered studies published only in the English language. No restrictions were applied to the year of publication, country of publication, comparator, type of publication, or study setting.

Study Selection

We followed 2 steps of the study selection process. In the first step, 2 reviewers (MA and AA) independently sifted the titles and abstracts of all retrieved studies. In the second step, the 2 reviewers independently scrutinized the full texts of the studies included in the first step. In both steps, any disagreements among the reviewers were resolved through discussion and consensus. Cohen κ in this review indicated a very good level of interrater agreement in the first (0.88) and second step (0.95) of the selection process [18].

Data Extraction

Multimedia Appendix 3 shows the data extraction form that was used in this review to precisely and systematically extract the data from the included studies. A total of 2 reviewers (MA and AA) independently conducted data extraction from the included studies, and they resolved any disagreements through discussion and consensus. Cohen κ showed a very good level of interrater agreement among the reviewers (0.85) [18].

Risk of Bias Assessment

To assess the risk of bias in the included studies, we used the Risk of Bias 2 tool, which is recommended by the Cochrane Collaboration [19]. This tool assesses RCTs in terms of five domains: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result [19]. Then, the overall risk of bias was determined for each study based on the risk of bias judgments in the five domains [19]. A total of 2 reviewers (MA and AA) independently assessed the risk of bias in the included studies, and any disagreements were resolved through discussion and consensus. Interrater agreement among the reviewers was very good (Cohen $\kappa=0.86$) [18]. We presented the results of the risk of bias assessment using a graph showing the reviewers' judgments about each *risk of bias* domain in the Results section. We also showed reviewers' judgments about each *risk of bias* domain for each included study using a figure in Multimedia Appendix 4 [10,20-24].

Data Synthesis

We synthesized the extracted data using narrative and statistical approaches. Specifically, meta-analysis was carried out when at least two studies assessed the same outcome of interest and reported sufficient data for the analysis (eg, mean difference, SD, and number of participants in each intervention group).

When the abovementioned conditions were not met, we narratively synthesized findings of the included studies. We grouped and synthesized the findings according to the measured outcomes (ie, PA, glycemic control, and weight change).

We conducted a meta-analysis using Review Manager 5.4, which is a software developed by Cochrane. We used the mean difference to assess the effect of each trial and the overall effect when the outcome data were continuous, and the outcome measure of each outcome was identical in the meta-analyzed studies. However, we used the standardized mean difference when, among studies, the outcome was measured using different tools. We selected a random effects model in the analysis because of the clinical heterogeneity among the meta-analyzed studies in terms of intervention characteristics (eg, its directionality, purpose, and frequency) and population characteristics (eg, sample size and mean age).

We assessed the clinical heterogeneity of the meta-analyzed studies by inspecting the characteristics of their interventions, outcomes, participants, and comparators. Further, we evaluated the statistical heterogeneity of the meta-analyzed studies. To do so, we calculated a chi-square P value and I^2 to evaluate the statistical significance of heterogeneity and degree of heterogeneity, respectively. We judged the meta-analyzed studies as heterogeneous when the chi-square P value was ≤ 0.05 [25]. The degree of heterogeneity was considered unimportant, moderate, substantial, or considerable when I^2 ranged from 0% to 40%, 30% to 60%, 50% to 90%, or 75% to 100%, respectively [25].

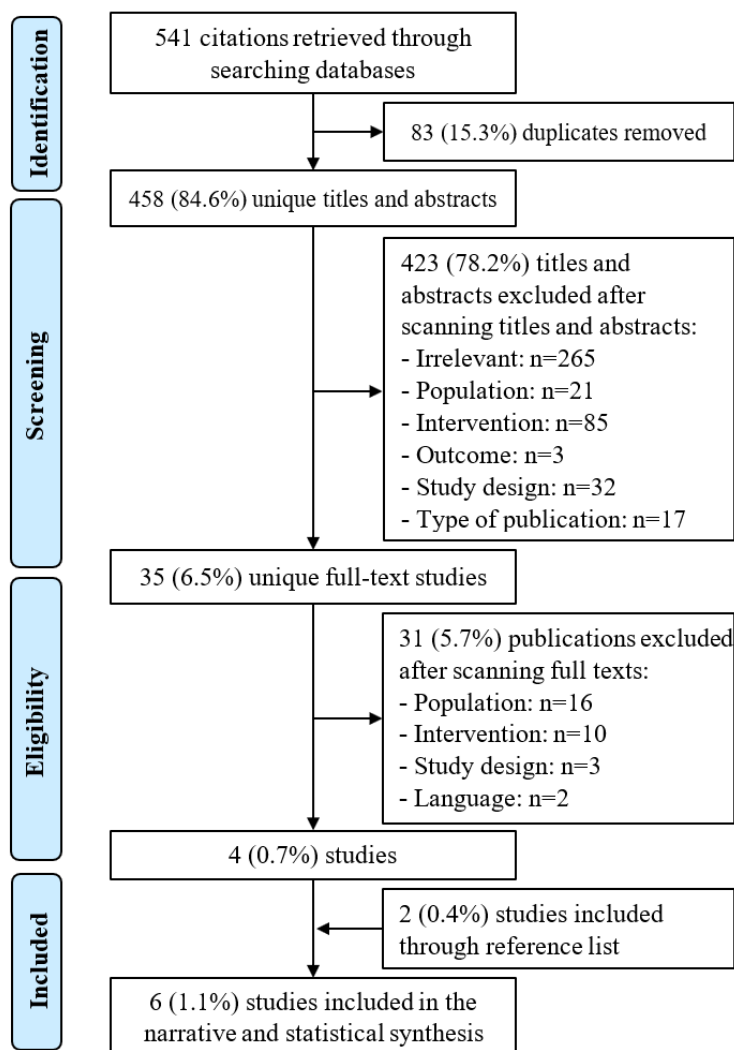
The overall quality of meta-analyzed evidence was examined using the Grading of Recommendations Assessment, Development, and Evaluation approach [26,27]. This approach assessed the quality of evidence based on five main criteria: risk of bias, inconsistency (ie, heterogeneity), indirectness, imprecision, and publication bias [26]. A total of 2 reviewers (MA and AA) independently assessed the overall quality of the meta-analyzed evidence, and any disagreements were resolved through discussion and consensus. Interrater agreement among the reviewers was very good (Cohen $\kappa=0.81$) [18].

Results

Search Results

We retrieved 541 citations by searching the 6 bibliographic databases (Figure 1). Of these 541 citations, 83 (15.3%) duplicates were identified and excluded. We screened the titles and abstracts of the remaining 84.6% (458/541) citations and excluded 78.2% (423/541) citations owing to reasons shown in Figure 1. By checking the full texts of the remaining 35 (6.5%) studies, 31 (5.7%) studies were not eligible for this review for several reasons (Figure 1). We identified 2 additional studies by backward reference list checking. Overall, we included 6 studies in this review [10,20-24]. At all steps, consensus was agreed between the 2 reviewers (MA and AA), and referral to a third reviewer was not required.

Figure 1. Flow chart of the study selection process.



Characteristics of Included Studies

As detailed in Table 1, all the included studies were RCTs. The included studies were conducted in 3 countries: the United States (n=3), Iran (n=2), and Indonesia (n=1); 4 of the studies were published in 2018. The sample size in the included studies ranged between 28 and 138, with an average of 81 (SD 40.03).

The mean age of participants in the included studies varied from 44.6 to 65.5 years, with an average of 51.6 years (SD 6.7). The percentage of men in the included studies ranged from 23.3% to 57.9%, with an average of 42.2% (SD 12.1). All studies recruited patients with T2DM. The included studies recruited participants from health care (n=5) and community (n=1).

Table 1. Characteristics of studies and population.

Study	Year	Country	Study design	Sample size	Age (years), mean (SD)	Sex (male)	Health condition	Setting
Agboola et al [20]	2016	United States	RCT ^a	126	51.4 (11.5)	48.4%	T2DM ^b	Health centers
Arovah et al [21]	2018	Indonesia	RCT	43	65.5 (5.8)	37.2%	T2DM	Public hospital
Lari et al [10]	2018	Iran	RCT	73	47.6 (9.1)	53.4%	T2DM	Diabetes clinics
Lari et al [22]	2018	Iran	RCT	76	48.2 (8.8)	57.9%	T2DM	Diabetes clinics
Polgreen et al [23]	2018	United States	RCT	138	44.6 (15.9)	23.3%	T2DM	Community
Ramirez and Wu [24]	2017	United States	RCT	28	52 (9.0)	33%	T2DM	Ambulatory care clinic

^aRCT: randomized controlled trial.

^bT2DM: type 2 diabetes mellitus.

The interventions in the included studies were text messages only (n=1), text messages and educational CD about PA (n=1), and text messages and pedometers (n=4; [Table 2](#)). Text messages were unidirectional (n=1), bidirectional (n=4), and both (ie, most messages were unidirectional, and some messages were bidirectional; n=1). The purpose of the text messages in the included studies was to educate participants about PA (n=4), remind them to wear the pedometer, review goals, or self-monitor and record their steps (n=4), provide them with feedback about their previous day's activity (n=3), motivate

them to walk and exercise more (n=2), and set step goals (n=1). The frequency of text messages sent to participants ranged between 2 per week and 3 per day. The intervention was delivered for 12 weeks in 4 studies and 24 weeks in 2 studies. The intervention in 5 studies was theoretically informed. Specifically, the following theories or models were used to develop the intervention: Social Cognitive Theory (n=2), Health Promotion Models (n=2), and Transtheoretical Model and Grounded Theory (n=1).

Table 2. Characteristics of interventions.

Study	Intervention	Directionality	Purpose	Frequency	Period	Theory used
Agboola et al [20]	SMS and pedometers	1- and 2-way	Education, motivation, reminder, and feedback	2/day	24 weeks	Transtheoretical model and grounded theory
Arovah et al [21]	SMS and pedometers	2-way	Motivation and reminder	1-3/day	12 weeks	Social Cognitive Theory
Lari et al [10]	SMS	2-way	Education	Phase 1: 2-3/day; phase 2: 2/week	Phase 1: 2 weeks; Phase 2: 10 weeks	Health promotion models
Lari et al [22]	SMS + educational CD	1-way	Education	2/week	12 weeks	Health promotion models
Polgreen et al [23]	Intervention 1: SMS text messaging (reminder) + SMS text messaging (goal setting) + pedometer; intervention 2: SMS text messaging (reminder)+pedometer	2-way	Reminders, feedback, and setting goals	Intervention 1: 2/day; intervention 2: 1/day	24 weeks	N/A ^a
Ramirez and Wu [24]	Intervention 1: SMS text messaging + pedometer	2-way	Education reminders and feedback	≥4/week	12 weeks	Social Cognitive Theory

^aN/A: not applicable.

The comparison group received pedometers in 4 of the studies or no intervention in 2 studies ([Table 3](#)). The pedometers were used by the participants for 12 weeks (n=2) or 24 weeks (n=2). The follow-up period ranged from 4 weeks to 24 weeks. The following outcomes of interest were assessed in the included

studies: PA (n=6), glycemic control indicators (n=3), weight (n=1), and BMI (n=1). Step count was the most common outcome measure used in the included studies (n=4), followed by HbA_{1c} (n=2), weight scale (n=2), and metabolic equivalent of task questionnaire (n=2).

Table 3. Characteristics of comparators and outcomes.

Study	Comparator	Period (week)	Follow-up (week)	Outcome	Outcome measure
Agboola et al [20]	Pedometers	24	24	PA ^a , glycemic control, and weight	Step count, weight scale, and HbA _{1c} ^b
Arovah et al [21]	Pedometers	12	12 and 24	PA and glycemic control	Step count, PAR ^c questionnaire, HbA _{1c} , fasting glucose, and 2-hour glucose
Lari et al [10]	No intervention	N/A ^d	4 and 12	PA	MET ^e questionnaire
Lari et al [22]	No intervention	N/A	4 and 12	PA	MET questionnaire
Polgreen et al [23]	Pedometers	24	12 and 24	PA and BMI	Step count, weight scale, and stadiometer
Ramirez and Wu [24]	Pedometers	12	6 and 12	PA	Step count

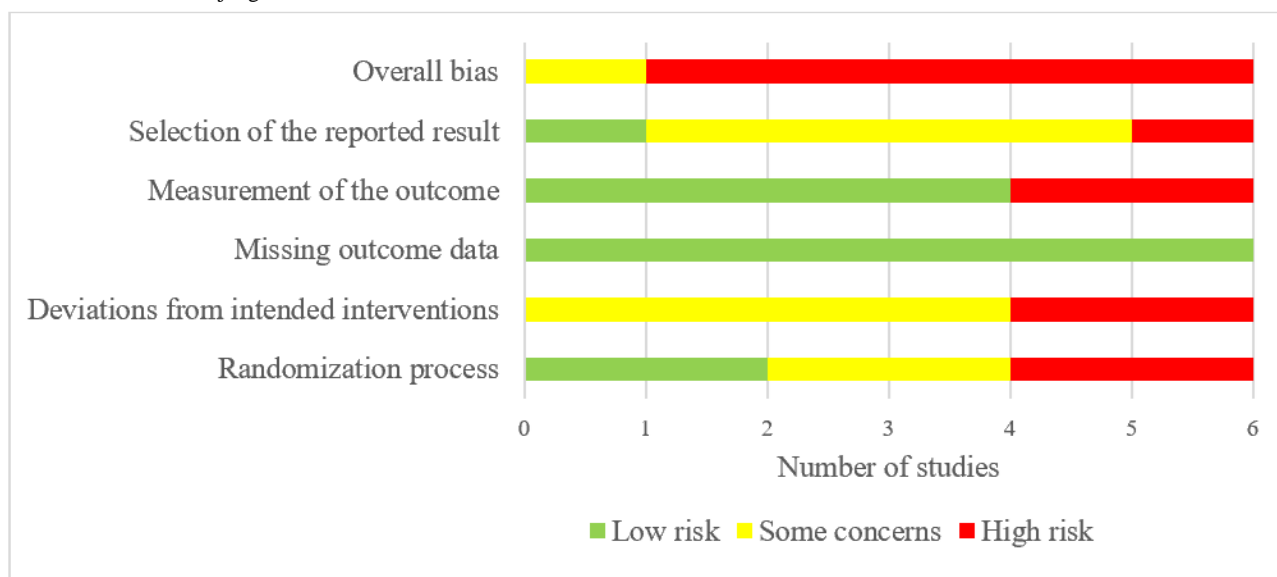
^aPA: physical activity.
^bHbA_{1c}: glycated hemoglobin.
^cPAR: physical activity rating.
^dN/A: not applicable.
^eMET: metabolic equivalent of task.

Risk of Bias Results

Although all studies used an appropriate random allocation sequence for the randomization process and had comparable groups, only 2 studies concealed the allocation sequence until participants were enrolled and assigned to interventions. Accordingly, only these 2 studies were rated as having a low risk of bias in the randomization process (Figure 2). In all

studies, participants, their health care professionals, researchers, or individuals delivering the interventions were aware of the assigned intervention during the trial. The study also did not report any information about whether a deviation from the intended intervention occurred owing to the experimental context. Thus, none of the studies were rated as having a low risk of bias in deviations from the intended interventions (Figure 2).

Figure 2. Review authors’ judgments about each risk of bias domain.



Outcome data were not available for all participants in the included studies, and there was no evidence that the findings were not biased by missing outcome data. However, the reasons for missing outcome data were not related to the true value of the outcome in all studies. Thus, all studies were judged as having a low risk of bias in the domain of missing outcome data.

In 4 studies, the outcomes of interest were assessed using appropriate measures (eg, pedometer and HbA_{1c}), which were comparable between the intervention groups. Therefore, these studies were rated as having a low risk of bias when measuring the outcome. However, the remaining 2 studies were judged as having a high risk of bias in this domain because they used subjective outcome measures that depended on participants’

recall, and participants and outcome assessors were not blinded in the 2 studies (Figure 2).

Only 1 study was judged as having a low risk of bias in the selection of the reported studies (Figure 2). This judgment is attributed to the fact that the remaining studies did not publish a prespecified analysis plan or reported outcome measurements and analyses different from those specified in the analysis plan. Given that 5 studies were judged as having a high risk of bias in at least one domain, they were rated as high risk in the domain of overall bias. The remaining study was judged to raise some concerns in the domain of overall bias, as it had some concerns in one of the domains. Reviewers' judgments about each risk of bias domain for each included study are presented in Multimedia Appendix 4.

Results of Studies

Effect on PA

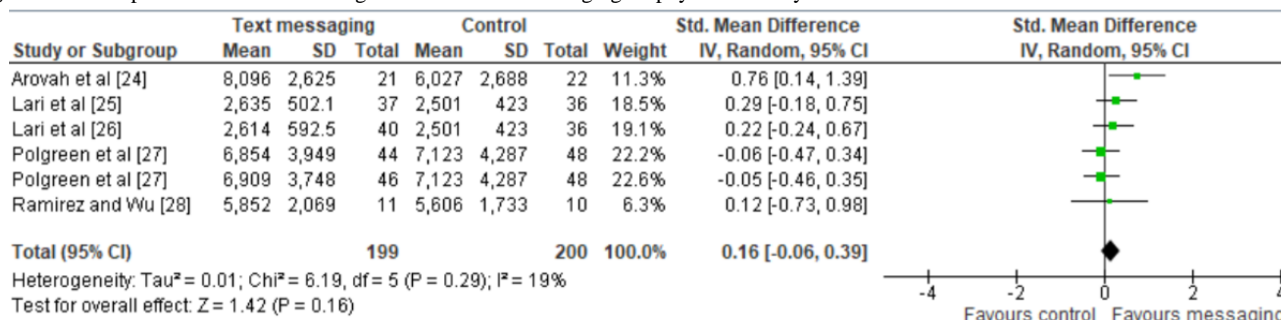
All included studies assessed the effect of using text messages on PA among patients with T2DM. A total of 3 studies showed a statistically significant effect of text messages on PA [10,21,22,24]. To be more precise, Arovah et al [21] compared the effect of text messages plus pedometers to only pedometers on PA as measured by daily step count, self-reported walking (min/week), and self-reported moderate-to-vigorous-intensity PA (min/week). The study showed a statistically significant effect of 12-week text messages plus pedometers to only pedometers on daily steps ($P<.001$), self-reported walking ($P=.001$), and moderate-to-vigorous-intensity PA ($P<.001$) [21]. In 2 further studies, where data were analyzed from different arms of a single RCT in each study, Lari et al [10] compared the effect of text messages only and text messages plus educational CD [22] to no intervention on PA as measured by the metabolic equivalent of task questionnaire. Both studies found a statistically significant effect of text messages only

($P<.001$) [10] and text messages plus educational CDs ($P<.001$) [22] on PA compared with no intervention.

The 3 remaining studies did not find a statistically significant effect of text messages on PA [20,23,24]. Specifically, Agboola et al [20] compared the effect of text messages plus pedometers to pedometers only on PA, as measured by the monthly step count. Although the study found that step counts over 6 months were higher in the intervention group than in the control group, this difference was not statistically significant ($P=.17$) [20]. Another study assessed the effect of text messages plus pedometers and only pedometers on PA, as assessed by daily steps [24]. The study did not show any statistically significant difference ($P=.78$) in PA between the 2 groups [24]. In a previous study, Polgreen et al [23] compared the effect of 2 interventions to only pedometers on PA, as measured by daily step count. The first intervention was pedometers plus text message reminders to wear the pedometers (reminders and pedometers), whereas the second intervention was the same as the first intervention plus text messages asking participants to set a step goal (goal setting, reminders, and pedometers) [23]. The study found no statistically significant differences in PA among the 3 groups [23].

A total of 5 studies were included in the statistical analysis (ie, meta-analysis), as they reported sufficient and appropriate data for the analysis [10,21-24]. The meta-analysis contained 6 comparisons as we included a comparison from each of the 4 studies [10,21,22,24] and 2 comparisons from the remaining study [23], which compared two types of text messages to no intervention. The meta-analysis showed no statistically significant difference in the PA ($P=.16$) between the text message group and the control group (standardized mean difference 0.16, 95% CI -0.06 to 0.39; Figure 3). The heterogeneity of the evidence was not a concern ($P=.29$; $I^2=19%$). The quality of the evidence was very low because of the high risk of bias and impression (Multimedia Appendix 5).

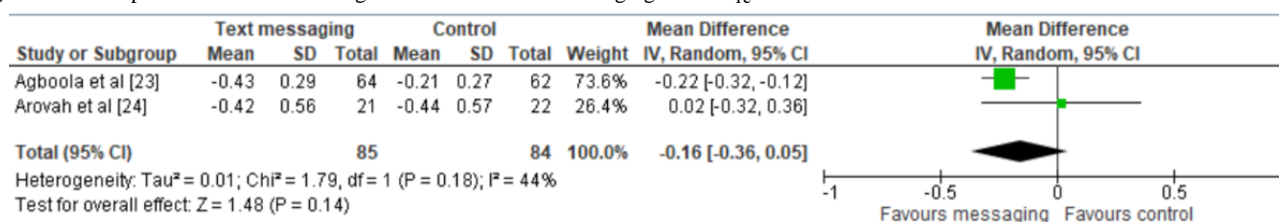
Figure 3. Forest plot of 6 studies assessing the effect of text messaging on physical activity.



Glycemic Control

A total of 2 studies examined the effect of text messages on glycemic control, as assessed by HbA_{1c} [20,21]. The results of both studies were meta-analyzed. The meta-analysis showed no statistically significant difference ($P=.14$) between the intervention and control groups, with no difference observed between text messages plus pedometers and only pedometers on HbA_{1c} (mean -0.16, 95% CI -0.36 to 0.05; Figure 4). There was moderate heterogeneity of the evidence ($I^2=44%$), but the

difference was not statistically significant ($P=.18$; Figure 4). The quality of evidence was low as it was downgraded by 1 level owing to a high risk of bias (Multimedia Appendix 5). It is worth mentioning that 1 of the 2 studies compared the effect of text messages plus pedometers with only pedometers on glycemic control as measured by fasting plasma glucose and 2-hour plasma glucose [21]. The study did not find a statistically significant difference between the groups in terms of fasting plasma glucose ($P=.18$) and 2-hour plasma glucose ($P=.90$) [21].

Figure 4. Forest plot of 2 studies assessing the effect of the text messaging on HbA_{1c}.

Anthropometric Measures

A total of 2 studies assessed anthropometric measures as outcomes (weight or BMI) [20,23]. The results of the 2 studies could not be statistically synthesized, as they assessed different outcomes. The first study showed no statistically significant difference between the intervention and control groups, with no effect of text messages plus pedometers on weight ($P=.77$) compared with pedometers alone [20]. In the second study, Polgreen et al [23] compared the effects of 2 interventions with only pedometers on BMI. The first intervention was pedometers plus text message reminders to wear the pedometers (reminders and pedometers), whereas the second intervention was the same as the first intervention plus text messages asking participants to set a step goal (goal setting, reminders, and pedometers) [23]. The study found no statistically significant differences in BMI among the 3 groups [23].

Other Outcomes

Secondary outcome measures reported in the examined studies included the following variables and parameters: reports of usability, satisfaction and adherence to the RCT as discussed in the study by Agboola et al [20], and quality of life or psychological outcomes (eg, self-efficacy, outcome expectations, self-regulation, and social support) as discussed in Arovah et al [21]. Lari et al [10,22] assessed the Health Promotion Model constructs (eg, perceived benefits, perceived barriers, perceived social support, and self-efficacy). Ramirez and Wu [24] also investigated the feasibility, perceived usefulness, and potential effectiveness.

Discussion

Principal Findings

This systematic review assessed the effectiveness of mobile text messaging as a method of promoting PA alone in people with T2DM. The meta-analysis of the results of 5 studies (6 comparisons) showed no statistically significant effect of mobile text messaging on PA in comparison with no intervention. The insignificant effect may be attributed to the fact that 3 studies showed a statistically significant effect of mobile text messaging on PA, whereas 2 studies did not find any significant effect of text messages on PA. There are several potential reasons for the significant increase in PA in 3 studies. First, the intervention in 1 study [21] was combined with pedometers, and some studies have found greater effects when using objective measures compared with subjective measures [28]. It is possible that participants in these studies were more active because of the knowledge that they were wearing the pedometer [29]. The remaining 2 RCTs [10,22] were rated as having a high risk of

bias because they used self-recall questionnaires to measure PA. However, these measures can present limitations in capturing PA because of poor reliability and validity, participant recall bias, and differences in the interpretation of questions [30]. Our findings are consistent with previous reviews that assessed the effect of text messaging on PA in participants with different chronic conditions [31]. Some studies observed only small improvements in daily steps and self-reported PA; other studies did not observe any statistically significant changes in PA despite the use of different PA measurement strategies [31].

Our review found no statistically significant effect of mobile text messaging on glycemic control as assessed by HbA_{1c}, fasting plasma glucose, and 2-hour plasma glucose. Our findings are consistent with those of previous studies that showed no significant difference in HbA_{1c} levels in people with T2DM following text messaging interventions [32]. This could be attributed to the duration effect, which had short interventions and follow-up durations (median 12 weeks); thus, outcomes such as HbA_{1c} are less likely to change over a short timescale (3 months). In other words, it might take longer for the intervention effects to become apparent [33].

The narrative synthesis in this review showed no statistically significant effect of mobile text messaging on either weight or BMI. We could not synthesize these measures in our meta-analysis because of the high heterogeneity in the included studies. Our findings are consistent with those of previous reviews, and a meta-analysis showed no statistically significant association between BMI and weight following mobile messaging interventions in people with T2DM [34]. However, it is important to be realistic about the period of intervention, and a longer period is required to determine the desired improvements in such clinical outcomes [35]. The aforementioned studies had short interventions (median 12 weeks); thus, outcomes such as weight and BMI are less likely to change on a short timescale [33].

Strengths and Limitations

Strengths

Our study is the first review and meta-analysis that focused on the effectiveness of text messages targeting only PA among T2DM patients. This enabled us to ensure that the effect of text messaging on PA outcomes is attributed to PA-related message content and to no other content such as diet, lifestyle, and general diabetes education. Our study is considered a robust and high-quality review given that we followed well-recommended guidelines (ie, PRISMA) in developing, executing, and reporting it.

To run as sensitive a search as possible, we searched the most popular databases in the health and information technology fields using a very comprehensive list of search terms. The risk of publication bias is minimal in this review because we searched gray literature databases (ie, Web of Science and Google Scholar) and conducted backward and forward reference list checking. We did not restrict our search to specific countries of publication, year of publication, comparators, or settings; thus, this resulted in a more comprehensive review.

The risk of selection bias was minimal in the current review as 2 authors (MA and AA) independently selected the studies, extracted data, and assessed the risk of bias and quality of evidence, and they had a very good interrater agreement in all processes. When possible, we meta-analyzed the results of the included studies, and this improved the power of studies and the estimates of the likely size of the effect of text messaging on different outcomes.

Limitations

The intervention of interest in this review was restricted to PA-related text messaging, so we did not examine the impact of other digital interventions, such as mobile apps, wearables, or other eHealth tools. We also focused on patients with T2DM, rather than patients with other types of diabetes. Accordingly, our results may not be generalizable to other eHealth interventions or patients with type 1 diabetes mellitus or gestational diabetes mellitus. In this review, we included only RCTs published in the English language; thus, it is possible that we missed results from some non-English RCTs. We applied these restrictions owing to the high internal validity of RCTs over other study designs [36] and lack of resources to translate non-English studies. The included studies were conducted in only 3 countries (the United States, Iran, and Indonesia); therefore, the generalizability of our findings to other countries may be limited. The findings were based on a small number of

studies that met the review criteria. Although 6 studies were included in this review, 2 (33%) of the studies were from a single RCT where 2 separate analyses were undertaken with data taken from different arms. Only 2 studies were included in each of the 2 meta-analyses conducted in this review. This is attributed to the lack of reported data that were appropriate for the analysis and incomparable outcome measures and comparators between studies. As such, it is not possible to draw firm conclusions about effectiveness.

Implications for Research

The current review found relatively few studies assessing the effectiveness of text messages in promoting PA in T2DM; thus, RCTs with larger sample sizes are needed. Future studies should seek to include objective outcome measures (eg, PA, glycemic control, and anthropometric measures), be consistent in terms of selected outcome measures, and measure outcomes after longer follow-up periods to be able to compare study findings and make firm conclusions about intervention effectiveness. More research is needed to determine the type of text message content, frequency of messaging, and duration of intervention that is most likely to result in positive outcomes. Additional research needs to include an estimation of the cost-effectiveness of text messages and an examination of their long-term impact.

Conclusions

We could not draw a definitive conclusion regarding the effectiveness of text messaging on PA, glycemic control, weight, or BMI among patients with T2MD, given the low number of included studies and their high risk of bias. Thus, the findings of this study suggest that texting messaging should not substitute but rather supplement clinical support. In addition, there is a pressing need for further RCTs with large sample sizes, low risk of bias, and more consistency regarding intervention duration, outcome measures, follow-up period, and comparator.

Acknowledgments

This review was supported by a doctoral scholarship from the Ministry of Higher Education, Saudi Arabia.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) checklist.

[[DOCX File, 23 KB - jmir_v24i3e29663_app1.docx](#)]

Multimedia Appendix 2

Search query used for searching MEDLINE.

[[DOCX File, 15 KB - jmir_v24i3e29663_app2.docx](#)]

Multimedia Appendix 3

Data extraction form.

[[DOCX File, 16 KB - jmir_v24i3e29663_app3.docx](#)]

Multimedia Appendix 4

Reviewers' judgements about each "risk of bias" domain for each included randomized controlled trial.

[[DOCX File , 64 KB - jmir_v24i3e29663_app4.docx](#)]

Multimedia Appendix 5

Grading of Recommendations Assessment, Development, and Evaluation profile.

[[DOCX File , 16 KB - jmir_v24i3e29663_app5.docx](#)]

References

1. Haider R, Sudini L, Chow CK, Cheung NW. Mobile phone text messaging in improving glycaemic control for patients with type 2 diabetes mellitus: a systematic review and meta-analysis. *Diabetes Res Clin Pract* 2019;150:27-37. [doi: [10.1016/j.diabres.2019.02.022](https://doi.org/10.1016/j.diabres.2019.02.022)] [Medline: [30822496](https://pubmed.ncbi.nlm.nih.gov/30822496/)]
2. Type 2 diabetes in adults: management. National Institute for Health and Care Excellence. 2019. URL: <https://www.nice.org.uk/guidance/ng28> [accessed 2020-03-05]
3. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2018. *Diabetes Care* 2018;41(Suppl 1):S13-S27. [doi: [10.2337/dc18-S002](https://doi.org/10.2337/dc18-S002)] [Medline: [29222373](https://pubmed.ncbi.nlm.nih.gov/29222373/)]
4. Gordon J, McEwan P, Idris I, Evans M, Puelles J. Treatment choice, medication adherence and glycemic efficacy in people with type 2 diabetes: a UK clinical practice database study. *BMJ Open Diabetes Res Care* 2018;6(1):e000512 [FREE Full text] [doi: [10.1136/bmjdc-2018-000512](https://doi.org/10.1136/bmjdc-2018-000512)] [Medline: [29755756](https://pubmed.ncbi.nlm.nih.gov/29755756/)]
5. Zheng Y, Ley SH, Hu FB. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat Rev Endocrinol* 2018;14(2):88-98. [doi: [10.1038/nrendo.2017.151](https://doi.org/10.1038/nrendo.2017.151)] [Medline: [29219149](https://pubmed.ncbi.nlm.nih.gov/29219149/)]
6. Edwards J, Hosseinzadeh H. The impact of structured physical activity on glycaemic control in diabetes prevention programmes: a systematic review. *Proc Singapore Healthc* 2017;27(3):193-204. [doi: [10.1177/2010105817739924](https://doi.org/10.1177/2010105817739924)]
7. Colberg SR, Sigal RJ, Yardley JE, Riddell MC, Dunstan DW, Dempsey PC, et al. Physical activity/exercise and diabetes: a position statement of the American diabetes association. *Diabetes Care* 2016;39(11):2065-2079 [FREE Full text] [doi: [10.2337/dc16-1728](https://doi.org/10.2337/dc16-1728)] [Medline: [27926890](https://pubmed.ncbi.nlm.nih.gov/27926890/)]
8. Diabetes Canada Clinical Practice Guidelines Expert Committee, Sigal RJ, Armstrong MJ, Bacon SL, Boulé NG, Dasgupta K, et al. Physical activity and diabetes. *Can J Diabetes* 2018;42 Suppl 1:S54-S63. [doi: [10.1016/j.cjcd.2017.10.008](https://doi.org/10.1016/j.cjcd.2017.10.008)] [Medline: [29650112](https://pubmed.ncbi.nlm.nih.gov/29650112/)]
9. Sibai AM, Costanian C, Tohme R, Assaad S, Hwalla N. Physical activity in adults with and without diabetes: from the 'high-risk' approach to the 'population-based' approach of prevention. *BMC Public Health* 2013;13:1002 [FREE Full text] [doi: [10.1186/1471-2458-13-1002](https://doi.org/10.1186/1471-2458-13-1002)] [Medline: [24153099](https://pubmed.ncbi.nlm.nih.gov/24153099/)]
10. Lari H, Noroozi A, Tahmasebi R. Impact of short message service (SMS) education based on a health promotion model on the physical activity of patients with type II diabetes. *Malays J Med Sci* 2018;25(3):67-77 [FREE Full text] [doi: [10.21315/mjms2018.25.3.7](https://doi.org/10.21315/mjms2018.25.3.7)] [Medline: [30899188](https://pubmed.ncbi.nlm.nih.gov/30899188/)]
11. Kannisto KA, Koivunen MH, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. *J Med Internet Res* 2014;16(10):e222 [FREE Full text] [doi: [10.2196/jmir.3442](https://doi.org/10.2196/jmir.3442)] [Medline: [25326646](https://pubmed.ncbi.nlm.nih.gov/25326646/)]
12. Sahin C, Courtney KL, Naylor PJ, E Rhodes R. Tailored mobile text messaging interventions targeting type 2 diabetes self-management: a systematic review and a meta-analysis. *Digit Health* 2019;5:2055207619845279 [FREE Full text] [doi: [10.1177/2055207619845279](https://doi.org/10.1177/2055207619845279)] [Medline: [31041110](https://pubmed.ncbi.nlm.nih.gov/31041110/)]
13. Thakkar J, Kurup R, Laba TL, 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;176(3):340-349. [doi: [10.1001/jamainternmed.2015.7667](https://doi.org/10.1001/jamainternmed.2015.7667)] [Medline: [26831740](https://pubmed.ncbi.nlm.nih.gov/26831740/)]
14. Berrouguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. *J Med Internet Res* 2016;18(6):e135 [FREE Full text] [doi: [10.2196/jmir.5066](https://doi.org/10.2196/jmir.5066)] [Medline: [27287668](https://pubmed.ncbi.nlm.nih.gov/27287668/)]
15. Saffari M, Ghanizadeh G, Koenig HG. Health education via mobile text messaging for glycemic control in adults with type 2 diabetes: a systematic review and meta-analysis. *Prim Care Diabetes* 2014;8(4):275-285. [doi: [10.1016/j.pcd.2014.03.004](https://doi.org/10.1016/j.pcd.2014.03.004)] [Medline: [24793589](https://pubmed.ncbi.nlm.nih.gov/24793589/)]
16. Connelly J, Kirk A, Masthoff J, MacRury S. The use of technology to promote physical activity in type 2 diabetes management: a systematic review. *Diabet Med* 2013;30(12):1420-1432. [doi: [10.1111/dme.12289](https://doi.org/10.1111/dme.12289)] [Medline: [23870009](https://pubmed.ncbi.nlm.nih.gov/23870009/)]
17. 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;339:b2700 [FREE Full text] [doi: [10.1136/bmj.b2700](https://doi.org/10.1136/bmj.b2700)] [Medline: [19622552](https://pubmed.ncbi.nlm.nih.gov/19622552/)]
18. Altman DG. *Practical statistics for medical research*. Boca Raton, FL: CRC press; 1990.
19. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019;366:14898. [doi: [10.1136/bmj.14898](https://doi.org/10.1136/bmj.14898)] [Medline: [31462531](https://pubmed.ncbi.nlm.nih.gov/31462531/)]

20. Agboola S, Jethwani K, Lopez L, Searl M, O'Keefe S, Kvedar J. Text to move: a randomized controlled trial of a text-messaging program to improve physical activity behaviors in patients with type 2 diabetes mellitus. *J Med Internet Res* 2016;18(11):e307 [FREE Full text] [doi: [10.2196/jmir.6439](https://doi.org/10.2196/jmir.6439)] [Medline: [27864165](https://pubmed.ncbi.nlm.nih.gov/27864165/)]
21. Arovah NI, Kushartanti BM, Washington TL, Heesch KC. Walking with Diabetes (WW-DIAB) programme a walking programme for Indonesian type 2 diabetes mellitus patients: a pilot randomised controlled trial. *SAGE Open Med* 2018;6:2050312118814391 [FREE Full text] [doi: [10.1177/2050312118814391](https://doi.org/10.1177/2050312118814391)] [Medline: [35154751](https://pubmed.ncbi.nlm.nih.gov/35154751/)]
22. Lari H, Tahmasebi R, Noroozi A. Effect of electronic education based on health promotion model on physical activity in diabetic patients. *Diabetes Metab Syndr* 2018;12(1):45-50. [doi: [10.1016/j.dsx.2017.08.013](https://doi.org/10.1016/j.dsx.2017.08.013)] [Medline: [28869152](https://pubmed.ncbi.nlm.nih.gov/28869152/)]
23. Polgreen LA, Anthony C, Carr L, Simmering JE, Evans NJ, Foster ED, et al. The effect of automated text messaging and goal setting on pedometer adherence and physical activity in patients with diabetes: a randomized controlled trial. *PLoS One* 2018;13(5):e0195797 [FREE Full text] [doi: [10.1371/journal.pone.0195797](https://doi.org/10.1371/journal.pone.0195797)] [Medline: [29718931](https://pubmed.ncbi.nlm.nih.gov/29718931/)]
24. Ramirez M, Wu S. Phone messaging to prompt physical activity and social support among low-income Latino patients with type 2 diabetes: a randomized pilot study. *JMIR Diabetes* 2017;2(1):e8 [FREE Full text] [doi: [10.2196/diabetes.7063](https://doi.org/10.2196/diabetes.7063)] [Medline: [30291094](https://pubmed.ncbi.nlm.nih.gov/30291094/)]
25. Deeks JJ, Higgins JP, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JP, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Sussex, UK: John Wiley & Sons; 2008:243-296.
26. Schunemann HJ, Oxman AD, Vist GE, Higgins JP, Deeks JJ, Glasziou P, et al. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JP, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Sussex, UK: John Wiley & Sons; 2008:359-387.
27. Walshaw MJ. Cystic fibrosis: diagnosis and management - NICE guideline 78. *Paediatr Respir Rev* 2019;31:12-14. [doi: [10.1016/j.prrv.2019.02.006](https://doi.org/10.1016/j.prrv.2019.02.006)] [Medline: [30962150](https://pubmed.ncbi.nlm.nih.gov/30962150/)]
28. Silfee VJ, Haughton CF, Jake-Schoffman DE, Lopez-Cepero A, May CN, Sreedhara M, et al. Objective measurement of physical activity outcomes in lifestyle interventions among adults: a systematic review. *Prev Med Rep* 2018;11:74-80 [FREE Full text] [doi: [10.1016/j.pmedr.2018.05.003](https://doi.org/10.1016/j.pmedr.2018.05.003)] [Medline: [29984142](https://pubmed.ncbi.nlm.nih.gov/29984142/)]
29. Bravata DM, Smith-Spangler C, Sundaram V, Gienger AL, Lin N, Lewis R, et al. Using pedometers to increase physical activity and improve health: a systematic review. *JAMA* 2007;298(19):2296-2304. [doi: [10.1001/jama.298.19.2296](https://doi.org/10.1001/jama.298.19.2296)] [Medline: [18029834](https://pubmed.ncbi.nlm.nih.gov/18029834/)]
30. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. *Prog Cardiovasc Dis* 2015;57(4):387-395. [doi: [10.1016/j.pcad.2014.10.005](https://doi.org/10.1016/j.pcad.2014.10.005)] [Medline: [25446555](https://pubmed.ncbi.nlm.nih.gov/25446555/)]
31. Monroe CM, Thompson DL, Bassett Jr DR, Fitzhugh EC, Raynor HA. Usability of mobile phones in physical activity-related research: a systematic review. *Am J Health Educ* 2015;46(4):196-206. [doi: [10.1080/19325037.2015.1044141](https://doi.org/10.1080/19325037.2015.1044141)]
32. Valles-Ortiz PM, Miranda-Felix P, García Sosa ES. Text messaging interventions to glycemic control in type 2 diabetes adults: systematic review. *Enfermeria Global* 2015;14(1):435-444.
33. Polonsky WH, Fisher L, Schikman CH, Hinnen DA, Parkin CG, Jelsovsky Z, et al. Structured self-monitoring of blood glucose significantly reduces A1C levels in poorly controlled, noninsulin-treated type 2 diabetes: results from the Structured Testing Program study. *Diabetes Care* 2011;34(2):262-267 [FREE Full text] [doi: [10.2337/dc10-1732](https://doi.org/10.2337/dc10-1732)] [Medline: [21270183](https://pubmed.ncbi.nlm.nih.gov/21270183/)]
34. Arambepola C, Ricci-Cabello I, Manikavasagam P, Roberts N, French DP, Farmer A. The impact of automated brief messages promoting lifestyle changes delivered via mobile devices to people with type 2 diabetes: a systematic literature review and meta-analysis of controlled trials. *J Med Internet Res* 2016;18(4):e86 [FREE Full text] [doi: [10.2196/jmir.5425](https://doi.org/10.2196/jmir.5425)] [Medline: [27095386](https://pubmed.ncbi.nlm.nih.gov/27095386/)]
35. Bin Abbas B, Al Fares A, Jabbari M, El Dali A, Al Orifi F. Effect of mobile phone short text messages on glycemic control in type 2 diabetes. *Int J Endocrinol Metab* 2015;13(1):e18791 [FREE Full text] [doi: [10.5812/ijem.18791](https://doi.org/10.5812/ijem.18791)] [Medline: [25745493](https://pubmed.ncbi.nlm.nih.gov/25745493/)]
36. Bhattacharjee A. *Social science research: Principles, methods, and practices*. Athens, GA: Global Text Project; 2012.

Abbreviations

HbA_{1c}: glycated hemoglobin

PA: physical activity

PRISMA: Preferred Reporting Item for Systematic Reviews and Meta-Analyses

RCT: randomized controlled trial

T2DM: type 2 diabetes mellitus

Edited by R Kukafka; submitted 15.04.21; peer-reviewed by E Al-Jafar, K Knittle; comments to author 03.05.21; revised version received 30.09.21; accepted 29.10.21; published 08.03.22.

Please cite as:

Alsahli M, Abd-Alrazaq A, Househ M, Konstantinidis S, Blake H

The Effectiveness of Mobile Phone Messaging–Based Interventions to Promote Physical Activity in Type 2 Diabetes Mellitus: Systematic Review and Meta-analysis

J Med Internet Res 2022;24(3):e29663

URL: <https://www.jmir.org/2022/3/e29663>

doi: [10.2196/29663](https://doi.org/10.2196/29663)

PMID: [35258463](https://pubmed.ncbi.nlm.nih.gov/35258463/)

©Mohammed Alsahli, Alaa Abd-Alrazaq, Mowafa Househ, Stathis Konstantinidis, Holly Blake. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 08.03.2022. 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.

Review

Standards, Processes, and Tools Used to Evaluate the Quality of Health Information Systems: Systematic Literature Review

René Noël^{1*}, MSc; Carla Taramasco^{1*}, PhD; Gastón Márquez^{2*}, PhD

¹Escuela de Ingeniería Informática, Universidad de Valparaíso, Valparaíso, Chile

²Departamento de Electrónica e Informática, Universidad Técnica Federico Santa María, Concepción, Chile

* all authors contributed equally

Corresponding Author:

Carla Taramasco, PhD

Escuela de Ingeniería Informática, Universidad de Valparaíso

General Cruz 222

Valparaíso, 2340000

Chile

Phone: 56 32 250 3630

Fax: 56 32 250 3630

Email: carla.taramasco@uv.cl

Abstract

Background: Evaluating health information system (HIS) quality is strategically advantageous for improving the quality of patient care. Nevertheless, few systematic studies have reported what methods, such as standards, processes, and tools, were proposed to evaluate HIS quality.

Objective: This study aimed to identify and discuss the existing literature that describes standards, processes, and tools used to evaluate HIS quality.

Methods: We conducted a systematic literature review using review guidelines focused on software and systems. We examined seven electronic databases—Scopus, ACM (Association for Computing Machinery), ScienceDirect, Google Scholar, IEEE Xplore, Web of Science, and PubMed—to search for and select primary studies.

Results: Out of 782 papers, we identified 17 (2.2%) primary studies. We found that most of the primary studies addressed quality evaluation from a management perspective. On the other hand, there was little explicit and pragmatic evidence on the processes and tools that allowed for the evaluation of HIS quality.

Conclusions: To promote quality evaluation of HISs, it is necessary to define mechanisms and methods that operationalize the standards in HISs. Additionally, it is necessary to create metrics that measure the quality of the most critical components and processes of HISs.

(*J Med Internet Res* 2022;24(3):e26577) doi:[10.2196/26577](https://doi.org/10.2196/26577)

KEYWORDS

health information systems; quality; standards; processes; metrics; systematic literature review

Introduction

The quality of information systems represents the set of qualities and properties that characterize and determine the usefulness and existence of these systems [1] (eg, security, usability, scalability, and others). Quality can be interpreted as a set of characteristics that a product or service possesses, as well as its capacity to satisfy new and complex user requirements (eg, security in medical records [2]). This implies that the product or service complies with the specifications for which it has been designed and must conform to those as expressed by users and

clients [3]. Some studies, such as Owens and Khazanchi [4], associate quality with (1) the explicitly stated functional and performance requirements, (2) the fully documented development standards, and (3) the implicit characteristics expected of any professionally developed system. On the other hand, the IEEE (Institute of Electrical and Electronics Engineers) defines quality as the degree to which a system, component, or process meets the specified requirements and the needs or expectations of the customer or user [5]. Both definitions denote that the emphasis of quality is on the specific requirements of the system and the pursuit of customer satisfaction.

Health information system (HIS) quality refers to whether a system's internal and external specifications and the expectations of stakeholders are satisfied [6]. The development of informatics and technology has enabled health professionals to work with large volumes of data and information, as well as to transmit them smoothly. In turn, information from HISs can be used to drive decision-making, policy, research, and, ultimately, health outcomes [7]. In this regard, the use of health information technology improves the quality and effectiveness of health care. Additionally, it promotes individual and public health and increases diagnostic accuracy [8].

HIS quality can be measured in several forms, where the leading indicators are those related to patient care and the system's components and structure [9]. Quality standards use methodologies for the design, programming, testing, and analysis of the developed system, with the objectives of offering (1) better reliability and maintainability that agree with the requirements demanded by users and (2) control of the quality of the system aiming at improving its effectiveness and efficiency [10]. In general, once the system has been validated as meeting the main functional requirements specified, the user will perform acceptance tests in order to deploy the system into the production environment.

HISs require methodologies and processes to evaluate their quality, since these systems map the diversity of health systems into explicit algorithmic functionalities, represented by software systems, which can inevitably produce problems in terms of efficiency or effectiveness in the work and daily activities of clinicians [11]. Traditionally, health services have been conceived as independent services where patients receive different types of medical care at different levels (ie, primary, secondary, and tertiary). This independence eventually leads to a lack of communication and coordination between services, which implies that the efficiency of an HIS is compromised [12]. Therefore, quality standards allow us to evaluate and standardize HISs in order to satisfy clinical requirements, define processes, reduce management problems, and develop HISs with high-quality standards. Although the range of quality standards in information systems, in general, is quite broad, there is little evidence of any compilation work that systematically identifies, evaluates, and describes evidence from primary studies related to quality standards in HISs. This situation makes it difficult to have a comprehensive perception of the most relevant aspects of the use and application of quality standards in HISs.

The importance and relevance of quality evaluation in HISs have been explored in several literature reviews. Villamor Ordozgoiti et al [13] conducted a literature review related to quality criteria in information and communication technologies in health care. The authors concluded that quality assessment that specifies health care systems' requirements, including management, clinical, diagnostic, or monitoring, should be oriented toward the perspective of the institutions as users, clients, and acquirers of software. Sousa and Lopez [14] addressed the problem of usability regarding health care systems and how this problem can compromise the quality of the systems. For this reason, the authors conducted a systematic review regarding the usability of eHealth tools. The review

results indicated that the poor usability of eHealth tools affects the possibility of adopting this type of system. Azad-Khaneghah et al [15] combined grey literature and academic literature reviews to evaluate mobile health (mHealth) apps' usability and quality. The authors noted that most of the current mobile app quality rating scales have not been developed for the general public. Nouri et al [16] conducted a systematic review addressing the quality assessment of mHealth apps. As a result of the study, the authors mentioned the enormous heterogeneity in the evaluation criteria of mHealth apps in different studies. This may be due to the researchers' various quality assessments or different definitions for each criterion. Triantafillou [17] conducted a narrative review on the quality management methods for electronic health records (EHRs). The results of this review indicated that there is substantial evidence that EHR systems contribute in various ways to improving quality management. Although there is a constant interest in quality assessment in several health care systems, to the best of our knowledge, few studies have explicitly addressed what standards, processes, and tools are used to assess HIS quality.

In this paper, we report the results of a systematic literature review (SLR) on methods used to evaluate the quality of HISs. The main objective of this review was to identify, characterize, and describe primary studies that complement the state-of-the-art of standards, processes, and tools used to evaluate HIS quality. We reviewed over 782 articles, from which we selected 17 (2.2%) primary studies. We analyzed, classified, and described each primary study in order to discuss the proposals for the quality evaluations of HISs. Based on our main objective, we defined the following research questions: What standards and certifications have been used to certify the quality of HISs? Which processes have been used to certify a software product? Which tools have been used by health software providers to certify a built software product?

The first research question addresses techniques and methods that allow us to describe patterns, models, or benchmarks to measure or evaluate quality of HISs. The second research question focuses mainly on describing phases or sets of activities that enable the evaluation of HIS quality. Finally, the third research question addresses technologies that facilitate the practical application of standards, certifications, and processes.

Methods

Identification

We used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [18] to conduct our SLR (Figure 1). Additionally, we used the PICO (population, intervention, comparison, and output) structure suggested by Petersen et al [19] to define a search string according to the population, intervention, comparison, and output. The following points describe the keywords for the PICO structure:

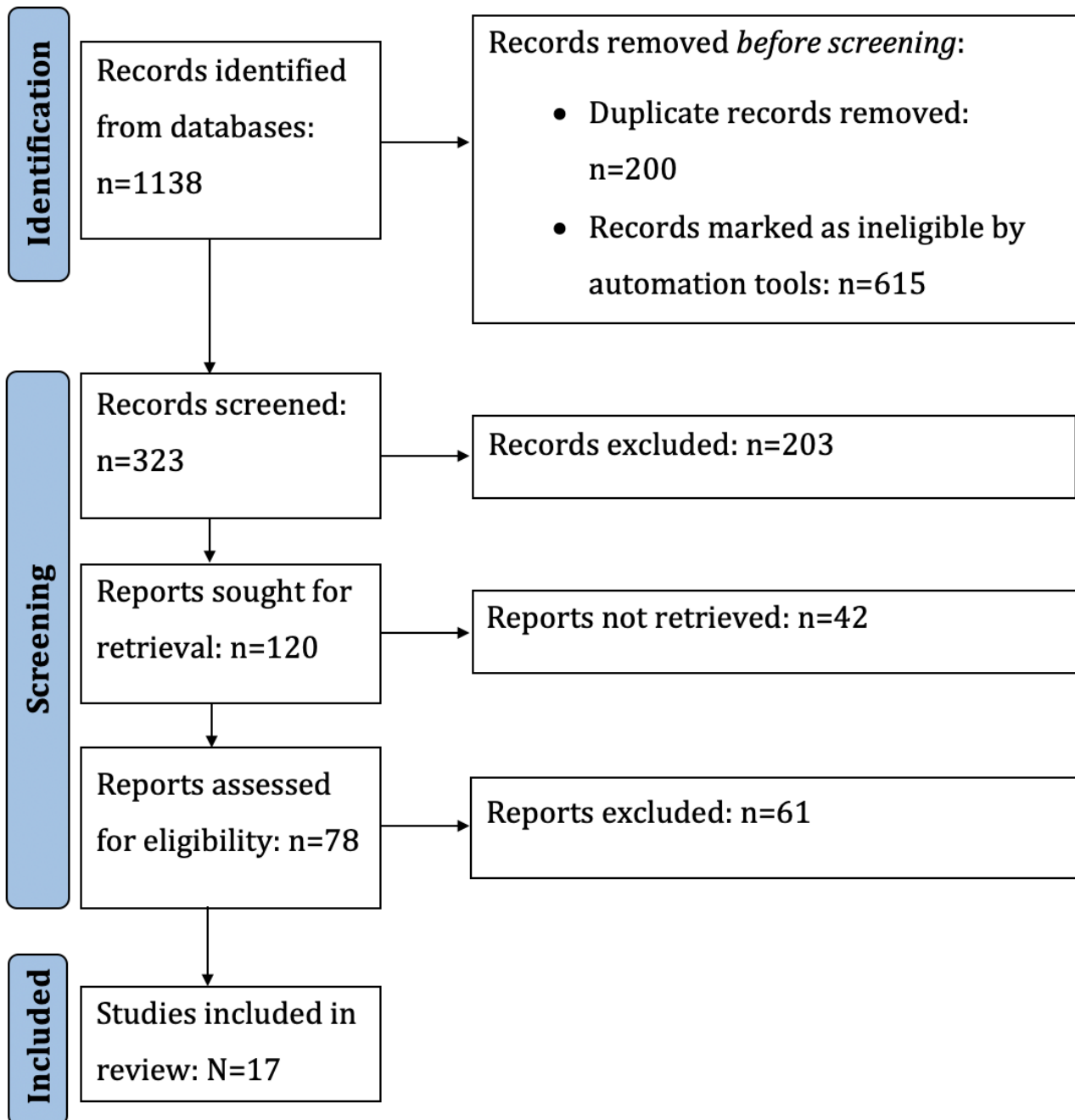
1. Population: articles related to HISs and synonyms ("health information system*" OR "e-health" OR "eHealth" OR "health software").
2. Intervention: articles related to the quality assessment of HISs and synonyms ("certification" OR "testing" OR

- “validation” OR “verification” OR “assessment” OR “legalization” OR “quality”).
3. Comparison: there were no previous studies on the subject that could be used as a baseline for comparison since our objectives and goals were not the same.
 4. Output: standards, tools, or processes applied in quality assessment (“process” OR “standard” OR “tool”).

system*” OR “e-health” OR “eHealth” OR “health software”) AND (“testing” OR “assessment” OR “validation” OR “verification” OR “certification”) AND (“process*” OR “standard*” OR “tool*”). We explored seven electronic databases—Scopus, ACM (Association for Computing Machinery), ScienceDirect, Google Scholar, IEEE Xplore, Web of Science, and PubMed—to search for and select primary studies.

We connected these concepts using “AND” and “OR” operators and obtained the following search string: (“health information

Figure 1. PRISMA flowchart of the selection of primary studies for the systematic literature review. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.



Screening

We screened primary studies using inclusion and exclusion criteria. We used the following inclusion criteria:

- The article is related to health
- The article provides HIS-related content and consolidated results from its research
- The article references or presents a standard, process, or tool
- The article addresses some quality attributes

- The article describes evaluation methods of HISs.

On the other hand, we applied the following exclusion criteria:

- The words “health” and “information system” are found, but they have no relation to our study
- The phrase “certification in health” is used, but its meaning is related to security
- The article is not related to standards, processes, or tools in the field of study
- The article is related to the quality attribute of interoperability.

Quality Assessment

The purpose of the quality evaluation was to evaluate the importance of each selected document. Although the quality assessment did not affect the selection of primary studies [20], we describe the evaluation primarily to reflect the selected studies' validity. According to each research question's answer, we evaluated each paper with 2, 1, or 0 points. Then, we chose those papers that exceeded the 50% threshold. The studies selected through this evaluation will ensure that our conclusions from the extracted data have some support from adequate resources (see [Multimedia Appendix 1](#) for more details on the quality assessment criteria).

Results

Overview

We identified 17 primary studies that were published in journals and in conference proceedings [21-37]. The scores of each selected primary study illustrate the quality and credibility of our study results. The average score was 74% (SD 0.11%), which means that the average quality of our study was acceptable (see [Multimedia Appendix 2](#) for more details on the quality assessment results).

The study publication years ranged from 2004 to 2020. We did not find studies that were published in 2005, 2006, 2008, 2011, 2012, or 2014. We did not find primary studies published in workshop proceedings or book chapters ([Multimedia Appendix 3](#)).

Out of 17 primary studies, 59% (n=10) corresponded to research conducted in Europe. On the other hand, 24% (n=4), 12% (n=2), and 6% (n=1) of the studies corresponded to research conducted in Asia, the Americas, and Oceania, respectively (see [Multimedia Appendix 4](#) for more detailed descriptions of the primary studies).

Standards and Certifications

The primary studies described several types of standards and certifications that are used in HISs ([Table 1](#)).

Table 1. Standards and certifications used in the primary studies.

Study No.	Standards and certifications	Reference
1	Usability heuristics	[22]
2	Telemedicine quality control	[23]
3	IHE ^a Connectathon, Q-REC ^b and ProRec ^c , CCHIT ^d , and others	[24]
4	The IHE initiative	[25]
5	Lean and agile principles	[26]
6	ISO ^e /IEC ^f 25010 standard	[27]
7	Custom certification framework	[28]
8	Custom usability principles applied in a case study	[29]
9	The uMARS ^g	[21]
10	The Constructive eHealth evaluation method	[30]
11	ISO 9241-210	[31]
12	The MDevSPICE framework	[32]
13	A care pathway data quality framework	[33]
14	The uMARS	[34]
15	The uMARS	[35]
16	The Medical Informatics Platform	[36]
17	The Medical Research Council framework	[37]

^aIHE: Integrating the Healthcare Enterprise.

^bQ-REC is a project entitled European Quality Labelling and Certification of Electronic Health Record Systems (EHRs).

^cThe ProRec initiative is a network of national nonprofit organizations (the “ProRec centres”) in Europe.

^dCCHIT: Certification Commission for Health Information Technology.

^eISO: International Organization for Standardization.

^fIEC: International Electrotechnical Commission.

^guMARS: end user version of the Mobile App Rating Scale.

A total of 65% (11/17) of the primary studies addressed standards that are related to quality management in HISs. Some of the primary studies, such as studies 1, 4, 5, and 10 (as numbered in Table 1), established that quality is based on the purposes and requirements established that must be met by any health care organization and the satisfaction of the needs of the people it serves. More precisely, HIS quality must allow for effective responses to health problems or situations that affect a population and its individuals, whether or not they identify them; HIS quality must also establish or apply the necessary diagnostic and therapeutic standards, procedures, and protocols to verify the medical instruments and means used. Studies 11 and 12 stated that translating quality policies to HISs can present diverse challenges that range from quality management to HIS implementation standards.

In addition, 35% (6/17) of the studies proposed metrics that allow for characterizing the standards. These metrics did not entirely represent a specific standard or certification, but rather they supported the evaluation of data control in telemedicine (study 2), the certification of EHRs (study 3), reliability in health-based mobile apps (studies 11, 15, and 17), and data management in medical platforms (study 16).

A total of 24% (4/17) of the primary studies addressed standards from a system perspective: 12% (2/17) addressed definition models, 6% (1/17) discussed design principles, and 6% (1/17) analyzed processes. In these studies, quality models and designs in HISs were discussed, but the implications of these proposals were not thoroughly discussed.

Processes

Generally, the primary studies discussed little information about processes that allow for certifying the quality of software products. Some studies, such as study 17, indicated that some of the reasons why there are no documented or described cases of processes to certify software are the costs of the processes.

Study 3 mentioned the processes defined by the ISO 9126 (International Organization for Standardization) standard to evaluate the quality of EHRs. This ISO standard evaluates all the characteristics of a software product from internal and external perspectives. Following the same perspective of quality in software products, study 6 mentioned that several scientific studies described a considerable increase in the number of users who surf the internet to obtain health-related information online. For this reason, information-seeking behavior on the web results in a need to ensure that web-based portals meet basic quality standards. Therefore, study 6 described the experience of

applying the ISO/IEC 25010 (International Electrotechnical Commission) quality assessment process to the e-Ebola Awareness System, an online health awareness portal. The process results provided some insights into the issues that negatively impacted the quality of the use of the portal, demanding attention and improvement.

A novel proposal, inspired by the ISO 9241-210 standard, was given by study 11. In this study, a human-centered design (HCD) approach was proposed to design connected health devices in order to ensure that user needs and requirements are considered throughout the design process. According to the authors, HCD is a multistage process that allows for several iterations of a design and subsequent updating of the requirements. Additionally, study 11 illustrated the implementation of an HCD by describing the techniques used to evaluate and develop usability and human factors in a case study addressing smartphone design and end user and stakeholder involvement.

Tools

A total of 29% (5/17) of the primary studies used well-known tools to certify the quality of health software. Studies 6 and 11 used the System Usability Scale (SUS) instrument [38]. This scale provides a fast and reliable tool for measuring usability. It consists of a 10-item questionnaire with five answer options for respondents, ranging from “strongly agree” to “strongly disagree.”

In addition to using the SUS, study 11 also used the After-Scenario Questionnaire (ASQ) instrument [39]. This questionnaire uses three statements to assess a user’s perceived difficulty with a task in a usability test. Studies 1, 10, and 15 addressed the Nielsen heuristics [40], which are 10 guidelines that measure usability through human-computer interaction. These heuristics aim to create systems that are as user friendly as possible.

Additionally, studies 1, 9, and 10 used other tools that were specified concisely. Study 9 proposed a tool called the end user version of the Mobile App Rating Scale (uMARS), which consists of reliability testing of a version of the Mobile App Rating Scale (MARS) for end users [21].

Discussion

Overview

Concerning standards (research question 1) and processes (research question 2), quality management and metrics concentrated the largest number of primary studies. The studies addressed quality as part of HIS management, which implies little detail on how quality standards were addressed in HISs. Additionally, there was no evidence about using processes that help manage HIS quality standards. In addition, four primary studies (studies 2, 15, 16, and 17) addressed different perspectives of metrics and did not discuss, in depth, what types of processes they used to apply the metrics to HISs. Other primary studies, such as studies 7, 9, and 13, addressed standards related to processes, design principles, and the definition of models but did not discuss the processes that support these standards.

Regarding processes (research question 2) and tools (research question 3), it is also worth noting the little discussion of these topics in the primary studies. Unlike study 9, which addressed a custom tool, only studies 3, 6, and 11 explicitly described the tools they used to evaluate HIS quality and also included them in the processes (ie, ISO/IEC 25010 and ISO 9126). However, a significant number of primary studies did not fully address quality assessment tools and supporting processes.

Another aspect described by some primary studies, such as studies 9, 14, 15, and 17, was that many clinicians are now taking advantage of the potential of mobile apps to address specific health problems. This implies that there must be tools to assess the appropriateness of usability regarding mobile apps. Some of these tools point to user acceptability, ease of use, and identification of risks in the use of mobile apps among patients.

Principal Findings

Our findings regarding HIS quality assessment revealed that there are several technical and social challenges to effectively achieving HIS quality objectives. More precisely, HIS quality assessment offers a method for evaluating the impact of changes in clinical processes that are embodied in systems. In general, the HIS represents an organized set of clinical functions involving people, data, activities, and overall material resources. These elements interact with each other to process data and information, including manual and automatic processes, in order to distribute them most appropriately within a given organization or entity based on its objectives.

Expanding the information on HIS quality requires the development of valid measurement instruments. The primary studies in this paper described some metrics, such as the SUS, Nielsen metrics, the ASQ, and others; however, these metrics only focus on one aspect of quality: usability. Usability in both HISs and the health sector is a critical attribute. Usability is defined as a measure of how well a specific user in a specific context can use a product or design to achieve a defined objective effectively, efficiently, and satisfactorily [40].

Other primary studies, such as studies 15 and 17, suggested that quality can be measured in how a user employs an HIS. In this regard, usability is one way to measure HIS quality. The primary studies mentioned several quality standards and processes, such as ISO 9126, ISO/IEC 62366:2015, and ISO/IEC 25010, that were used to evaluate HIS quality. These standards are composed of multidimensional attributes. For example, the ISO/IEC 25010 standard considers eight categories, as follows: (1) functional adequacy, (2) performance efficiency, (3) compatibility, (4) usability, (5) reliability, (6) security, (7) maintainability, and (8) portability. In turn, each category of the standard is divided into more quality attributes. Considering this example, it is natural to ask how HIS quality can be evaluated using the ISO/IEC 25010 standard as a reference. Extending this question to a more general scenario, another concern that emerges is what quality attributes are relevant in HISs.

The primary studies provided procedures and methods for evaluating HIS quality. Nevertheless, one aspect that we noticed is that there is no precise description of how to translate these

standards into practice. Taking the example of the ISO/IEC 25010 standard, there is a considerable set of quality attributes that allow quality to be established in HISs. However, there is little information regarding success stories, case studies, or other empirical studies that describe the lessons learned about applying quality standards in HISs. This lack of information does not allow for the replication of results in other HISs in order to build a body of knowledge related to HIS quality. In addition, the primary studies described very discreetly what lessons they learned from applying quality standards to HISs. HISs involve not only technical aspects but also social aspects. The quality standards address the technical aspects of HIS quality but leave the social aspects of HIS quality to be addressed.

In the Results section, we described the metrics that were reported in the primary studies. However, these metrics addressed general aspects of information system usability. Although these metrics greatly contributed to evaluating HIS usability, they did not address other clinical aspects relevant to clinicians. Some metrics, such as NISTIR 7804 (National Institute of Standards and Technology Interagency/Internal Report) and Health-ITUES (Health Information Technology Usability Evaluation Scale), measure usability in specific HISs (eg, EHRs), but again, these studies fell short of measuring usability. Therefore, to expand the boundaries of quality evaluation, it is necessary to identify which quality attributes are most relevant to evaluating HIS quality. Once these attributes have been identified, it is possible to conduct research on defining precise metrics for evaluating HIS quality. Institutions such as the World Health Organization have proposed various tools to evaluate different aspects of HISs, such as organization, clinical staff, technologies, and others.

The findings identified in this review provide a first impression of the emerging challenges involved in HIS quality assessment. HISs are complex and, as such, considerable effort is required to evaluate the quality of a complex system. There is no doubt that more than one metric should be proposed to evaluate all HIS components, whether social or technical. However, it is also desirable to share the lessons learned regarding conducting HIS quality assessments. In this way, the results can be replicated to help health care institutions evaluate their HISs in order to improve the quality of care for their patients.

Additionally, our review revealed lines of research attempting to increase the body of knowledge on HIS quality assessments. Challenges related to identifying, describing, and characterizing relevant quality attributes for assessing HIS quality; creating and validating accurate quality metrics and instruments; and reporting success stories and empirical evidence regarding HIS quality assessments are just some possible research challenges that can be addressed by the community. Furthermore, the creation of a multidimensional metric to assess quality in HISs is also considered a challenge. Given that HISs are composed of several different components, proposing a metric that assesses quality in a cross-dimensional way requires further research.

The benefits of using processes to evaluate and certify HIS quality are positive. According to Love and Li [41], one of the main benefits is improving clinical process efficiency and

effectiveness. This implies that the clinical services offered to patients are of good quality. From an internal management point of view, the processes help improve the internal communication capacity (ie, clinical services) and allow the different departments of a clinical institution to work together to satisfy patients' needs and expectations. However, our study's results revealed that the primary studies did not discuss, in depth, the use of processes to certify HISs. On this point, Love and Li [41] mentioned that applying quality certification processes can be highly demanding. For example, the costs of applying an HIS quality evaluation process are high, so such evaluations are often limited to organizations that have the resources for the evaluation. In addition, a large number of professionals are needed to conduct these assessments. This implies that the organization must have the resources to hire professionals and successfully conduct the assessment process. Under these scenarios, since health care institutions allocate their resources mainly to clinical and in-hospital management in order to care for patients, HIS evaluation does not necessarily rank high in managers' priorities.

Limitations

We critically reviewed the threats to the validity of our study. Because we conducted an SLR, this study suffers from the possible incompleteness of the search results and general publication bias. To analyze the threats to the validity of our study, we used the classification from Wohlin [42], which describes guidelines for classifying and mitigating threats to validity.

The threats to internal validity correspond to the factors that can affect the results of the study. In this regard, the bias in study selection is related to the potential bias in the search for articles in SLRs. To mitigate this threat, we used a robust literature review process on systems and software [20]. Additionally, we defined strict inclusion and exclusion criteria to select the primary studies. On the other hand, we are aware that the sample of studies obtained in this review was low ($N=17$). However, we performed several cross-check validations with three external collaborators to validate and evaluate each primary study. It is important to note that this SLR focused on HIS quality and not health quality. Therefore, after evaluating more than 700 studies, we determined that 17 primary studies satisfied our quality standards.

The threats to external validity are related to the restrictions that allow for the generalization of the results. The main threat is whether the primary studies represent HIS quality standards. To mitigate this threat, we invited health professionals from the Chilean National Center for Health Information Systems (CENS) to discuss and analyze each primary study in order to obtain feedback.

Regarding the threats to construction validity, these threats correspond to the generalization of the results to the concept underlying the execution of the SLR. The main threat is the subjectivity of our results. To mitigate this threat, we invited three external collaborators to support us in executing the SLR's main steps and comparing the results independently.

Conclusions

In this paper, we described the results of an SLR on HIS quality standards. We defined a rigorous process for identifying, characterizing, and evaluating this academic literature. As a result, we obtained 17 primary studies.

We identified five categories to classify standards. These categories are the definition of models, design principles, metrics, management, and processes. Most of the primary studies concentrated on the management category. We also realized that there is little information regarding processes that can be used to evaluate HIS quality. Some primary studies described evidence of ISO standards' evaluation processes, such as ISO 9126 and ISO/IEC 25010, but most did not detail their information. Finally, the primary studies were not clear in explicitly describing what tools they used to certify HIS quality.

The evidence for tools found in the primary studies suggests that these tools were usability evaluations.

Our findings point to primary studies agreeing that evaluating HIS quality is a relevant aspect of providing quality care to patients. However, several challenges compromise quality assessment. These challenges correspond to (1) the operationalization of quality certifications and standards in HISs, (2) the poor description of the metrics to measure HIS quality, and (3) the high demand for resources to conduct HIS quality assessments. To address these challenges, health care institutions must understand the importance of constantly assessing HIS quality. Defining standards, metrics, and processes for assessing quality provides countless benefits for HISs and contributes to creating quality-of-care models that ensure that patients receive appropriate treatments, thus minimizing the probability of errors.

Acknowledgments

We thank Paulina Silva, Alvaro Leiva, and Francisco Vera for their contributions to this study. Additionally, this research is supported by the Regular FONDECYT (Fondo Nacional de Desarrollo Científico y Tecnológico) via the project "Multimodal machine learning approach for the detection of pathological activity patterns in the elderly" (project No. 1201787). We also thank CENS and CORFO (Corporación de Fomento de la Producción) for support (project No. 16CTTS-66390) as well as FONDEF (Fondo de Fomento al Desarrollo Científico y Tecnológico) and IDeA (Investigación y Desarrollo en Acción) for support of the project "Sistema inteligente para la gestión y análisis de la dotación de camas en la red asistencial del sector público" (project No. ID16I10449).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Quality assessment criteria and the assignment of scores.

[[DOCX File, 15 KB - jmir_v24i3e26577_app1.docx](#)]

Multimedia Appendix 2

Quality assessment results.

[[DOCX File, 116 KB - jmir_v24i3e26577_app2.docx](#)]

Multimedia Appendix 3

Primary study distribution over the years. Numbers inside the circles represent the number of primary studies.

[[DOCX File, 68 KB - jmir_v24i3e26577_app3.docx](#)]

Multimedia Appendix 4

Descriptions of the primary studies.

[[DOCX File, 19 KB - jmir_v24i3e26577_app4.docx](#)]

References

1. Checkland P, Holwell S. Information, Systems and Information Systems: Making Sense of the Field. Chichester, UK: John Wiley & Sons Ltd; 1998.
2. Durneva P, Cousins K, Chen M. The current state of research, challenges, and future research directions of blockchain technology in patient care: Systematic review. *J Med Internet Res* 2020 Jul 20;22(7):e18619 [[FREE Full text](#)] [doi: [10.2196/18619](https://doi.org/10.2196/18619)] [Medline: [32706668](https://pubmed.ncbi.nlm.nih.gov/32706668/)]
3. Bates DW, Pappius E, Kuperman GJ, Sittig D, Burstin H, Fairchild D, et al. Using information systems to measure and improve quality. *Int J Med Inform* 1999 Feb;53(2-3):115-124. [doi: [10.1016/s1386-5056\(98\)00152-x](https://doi.org/10.1016/s1386-5056(98)00152-x)]
4. Owens DM, Khazanchi D. Software quality assurance. In: *Handbook of Research on Technology Project Management, Planning, and Operations*. Hershey, PA: IGI Global; 2009:242-260.

5. IEEE Standard for Software Quality Assurance Processes. In: IEEE Standards 730-2014. Piscataway, NJ: IEEE Standards Association; Jun 13, 2014:1-138.
6. 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](https://doi.org/10.3390/ijerph110505170)] [Medline: [24830450](https://pubmed.ncbi.nlm.nih.gov/24830450/)]
7. Mulgund P, Sharman R, Anand P, Shekhar S, Karadi P. Data quality issues with physician-rating websites: Systematic review. *J Med Internet Res* 2020 Sep 28;22(9):e15916 [FREE Full text] [doi: [10.2196/15916](https://doi.org/10.2196/15916)] [Medline: [32986000](https://pubmed.ncbi.nlm.nih.gov/32986000/)]
8. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: Impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006 May 16;144(10):742-752 [FREE Full text] [doi: [10.7326/0003-4819-144-10-200605160-00125](https://doi.org/10.7326/0003-4819-144-10-200605160-00125)] [Medline: [16702590](https://pubmed.ncbi.nlm.nih.gov/16702590/)]
9. Davis K, Doty MM, Shea K, Stremikis K. Health information technology and physician perceptions of quality of care and satisfaction. *Health Policy* 2009 May;90(2-3):239-246. [doi: [10.1016/j.healthpol.2008.10.002](https://doi.org/10.1016/j.healthpol.2008.10.002)] [Medline: [19038472](https://pubmed.ncbi.nlm.nih.gov/19038472/)]
10. Losavio F, Chirinos L, Matteo A, Lévy N, Ramdane-Cherif A. ISO quality standards for measuring architectures. *J Syst Softw* 2004 Jul;72(2):209-223. [doi: [10.1016/s0164-1212\(03\)00114-6](https://doi.org/10.1016/s0164-1212(03)00114-6)]
11. Bowman S. Impact of electronic health record systems on information integrity: Quality and safety implications. *Perspect Health Inf Manag* 2013;10:1c [FREE Full text] [Medline: [24159271](https://pubmed.ncbi.nlm.nih.gov/24159271/)]
12. Lega F, Prenestini A, Spurgeon P. Is management essential to improving the performance and sustainability of health care systems and organizations? A systematic review and a roadmap for future studies. *Value Health* 2013;16(1 Suppl):S46-S51 [FREE Full text] [doi: [10.1016/j.jval.2012.10.004](https://doi.org/10.1016/j.jval.2012.10.004)] [Medline: [23317645](https://pubmed.ncbi.nlm.nih.gov/23317645/)]
13. Villamor Ordozgoiti A, Delgado Hito P, Guix Comellas EM, Fernandez Sanchez CM, Garcia Hernandez M, Lluch Canut T. Software quality evaluation models applicable in health information and communications technologies. A review of the literature. *Stud Health Technol Inform* 2016;226:169-172. [Medline: [27350495](https://pubmed.ncbi.nlm.nih.gov/27350495/)]
14. Sousa V, Dunn Lopez K. Towards usable e-Health. A systematic review of usability questionnaires. *Appl Clin Inform* 2017 May 10;8(2):470-490 [FREE Full text] [doi: [10.4338/ACI-2016-10-R-0170](https://doi.org/10.4338/ACI-2016-10-R-0170)] [Medline: [28487932](https://pubmed.ncbi.nlm.nih.gov/28487932/)]
15. Azad-Khaneghah P, Neubauer N, Miguel Cruz A, Liu L. Mobile health app usability and quality rating scales: A systematic review. *Disabil Rehabil Assist Technol* 2021 Oct;16(7):712-721. [doi: [10.1080/17483107.2019.1701103](https://doi.org/10.1080/17483107.2019.1701103)] [Medline: [31910687](https://pubmed.ncbi.nlm.nih.gov/31910687/)]
16. Nouri R, Niakan Kalhori SR, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: A systematic review. *J Am Med Inform Assoc* 2018 Aug 01;25(8):1089-1098 [FREE Full text] [doi: [10.1093/jamia/ocy050](https://doi.org/10.1093/jamia/ocy050)] [Medline: [29788283](https://pubmed.ncbi.nlm.nih.gov/29788283/)]
17. Triantafyllou P. Making electronic health records support quality management: A narrative review. *Int J Med Inform* 2017 Aug;104:105-119. [doi: [10.1016/j.ijmedinf.2017.03.003](https://doi.org/10.1016/j.ijmedinf.2017.03.003)] [Medline: [28599812](https://pubmed.ncbi.nlm.nih.gov/28599812/)]
18. 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](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
19. Petersen K, Vakkalanka S, Kuzniarz L. Guidelines for conducting systematic mapping studies in software engineering: An update. *Inf Softw Technol* 2015 Aug;64:1-18. [doi: [10.1016/j.infsof.2015.03.007](https://doi.org/10.1016/j.infsof.2015.03.007)]
20. Kitchenham B, Charters S. Guidelines for Performing Systematic Literature Reviews in Software Engineering. Version 2.3. EBSE Technical Report. EBSE-2007-01. 2007 Jul 09. URL: https://www.elsevier.com/_data/promis_misc/525444systematicreviewsguide.pdf [accessed 2022-02-14]
21. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). *JMIR Mhealth Uhealth* 2016 Jun 10;4(2):e72 [FREE Full text] [doi: [10.2196/mhealth.5849](https://doi.org/10.2196/mhealth.5849)] [Medline: [27287964](https://pubmed.ncbi.nlm.nih.gov/27287964/)]
22. Kushniruk AW, Patel VL. 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](https://doi.org/10.1016/j.jbi.2004.01.003)] [Medline: [15016386](https://pubmed.ncbi.nlm.nih.gov/15016386/)]
23. Giansanti D, Morelli S, Macellari V. Experience at Italian National Institute of Health in the assessment of the quality in telemedicine: Submission to a quality control process. In: Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2007 Presented at: 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; August 22-26, 2007; Lyon, France p. 3725-3728. [doi: [10.1109/IEMBS.2007.4353141](https://doi.org/10.1109/IEMBS.2007.4353141)]
24. Hoerbst A, Schabetsberger T, Hackl W, Ammenwerth E. Requirements regarding quality certification of electronic health records. In: Proceedings of the XXIIInd International Congress of the European Federation for Medical Informatics. 2009 Presented at: XXIIInd International Congress of the European Federation for Medical Informatics; August 30-September 2, 2009; Sarajevo, Bosnia p. 384-388.
25. Virkanen H, Mykkänen J, Kajaste T. Status of interoperability requirements related to IHE integration profiles in Finland. *Finn J eHealth eWelfare* 2010;2(2):70-78 [FREE Full text]
26. Manjunath K, Jagadeesh N, Yogeesh M. Achieving quality product in a long term software product development in healthcare application using lean and agile principles: Software engineering and software development. In: International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing. 2013 Presented at:

- International Mutli-Conference on Automation, Computing, Communication, Control and Compressed Sensing; March 22-23, 2013; Kottayam, India p. 26-34. [doi: [10.1109/iMac4s.2013.6526379](https://doi.org/10.1109/iMac4s.2013.6526379)]
27. Hussain A, Mkpojiogu EO. An application of the ISO/IEC 25010 standard in the quality-in-use assessment of an online health awareness system. *Jurnal Teknologi* 2015 Nov 05;77(5):9-13. [doi: [10.11113/jt.v77.6107](https://doi.org/10.11113/jt.v77.6107)]
 28. Benedict M, Burwitz M, Schlieter H. Certification of service-oriented eHealth platform. In: *Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies*. 2015 Presented at: International Joint Conference on Biomedical Engineering Systems and Technologies; January 12-15, 2015; Lisbon, Portugal p. 114-122. [doi: [10.5220/0005221401140122](https://doi.org/10.5220/0005221401140122)]
 29. Unertl K, Holden RJ, Lorenzi NM. Usability: Making it real from concepts to implementation and end-user adoption. In: Weaver C, Ball M, Kim G, Kiel J, editors. *Healthcare Information Management Systems*. 4th Edition. Cham, Switzerland: Springer; 2016:165-175.
 30. Høstgaard AMB, Bertelsen P, Nøhr C. Constructive eHealth evaluation: Lessons from evaluation of EHR development in 4 Danish hospitals. *BMC Med Inform Decis Mak* 2017 Apr 20;17(1):45-15 [FREE Full text] [doi: [10.1186/s12911-017-0444-2](https://doi.org/10.1186/s12911-017-0444-2)] [Medline: [28427407](https://pubmed.ncbi.nlm.nih.gov/28427407/)]
 31. Harte R, Quinlan LR, Glynn L, Rodríguez-Molinero A, Baker PM, Scharf T, et al. Human-centered design study: Enhancing the usability of a mobile phone app in an integrated falls risk detection system for use by older adult users. *JMIR Mhealth Uhealth* 2017 May 30;5(5):e71 [FREE Full text] [doi: [10.2196/mhealth.7046](https://doi.org/10.2196/mhealth.7046)] [Medline: [28559227](https://pubmed.ncbi.nlm.nih.gov/28559227/)]
 32. Özcan-Top Ö, McCaffery F. A hybrid assessment approach for medical device software development companies. *J Softw* 2017 Dec 28;30(7):e1929. [doi: [10.1002/smr.1929](https://doi.org/10.1002/smr.1929)]
 33. Fox F, Aggarwal VR, Whelton H, Johnson O. A data quality framework for process mining of electronic health record data. In: *Proceedings of the IEEE International Conference on Healthcare Informatics*. 2018 Presented at: IEEE International Conference on Healthcare Informatics; June 4-7, 2018; New York, NY p. 12-21. [doi: [10.1109/ICHI.2018.00009](https://doi.org/10.1109/ICHI.2018.00009)]
 34. Li Y, Ding J, Wang Y, Tang C, Zhang P. Nutrition-related mobile apps in the China app store: Assessment of functionality and quality. *JMIR Mhealth Uhealth* 2019 Jul 30;7(7):e13261 [FREE Full text] [doi: [10.2196/13261](https://doi.org/10.2196/13261)] [Medline: [31364606](https://pubmed.ncbi.nlm.nih.gov/31364606/)]
 35. Bardus M, Ali A, Demachkieh F, Hamadeh G. Assessing the quality of mobile phone apps for weight management: User-centered study with employees from a Lebanese university. *JMIR Mhealth Uhealth* 2019 Jan 23;7(1):e9836 [FREE Full text] [doi: [10.2196/mhealth.9836](https://doi.org/10.2196/mhealth.9836)] [Medline: [30672742](https://pubmed.ncbi.nlm.nih.gov/30672742/)]
 36. Demiraj A, Karozos K, Spartalis I, Vassalos V. Meta-data management and quality control for the medical informatics platform. In: *Proceedings of the 23rd International Database Applications & Engineering Symposium*. 2019 Presented at: 23rd International Database Applications & Engineering Symposium; June 10-12, 2019; Athens, Greece p. 1-9. [doi: [10.1145/3331076.3331088](https://doi.org/10.1145/3331076.3331088)]
 37. Hawley-Hague H, Tacconi C, Mellone S, Martinez E, Ford C, Chiari L, et al. Smartphone apps to support falls rehabilitation exercise: App development and usability and acceptability study. *JMIR Mhealth Uhealth* 2020 Sep 28;8(9):e15460 [FREE Full text] [doi: [10.2196/15460](https://doi.org/10.2196/15460)] [Medline: [32985992](https://pubmed.ncbi.nlm.nih.gov/32985992/)]
 38. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the System Usability Scale. *Int J Hum Comput Interact* 2008 Jul 30;24(6):574-594. [doi: [10.1080/10447310802205776](https://doi.org/10.1080/10447310802205776)]
 39. Lewis J. Psychometric evaluation of an after-scenario questionnaire for computer usability studies. *SIGCHI Bull* 1991 Jan;23(1):78-81. [doi: [10.1145/122672.122692](https://doi.org/10.1145/122672.122692)]
 40. Nielsen J. Usability inspection methods. In: *Proceedings of the ACM Conference on Human Factors in Computer Systems*. 1994 Presented at: ACM Conference on Human Factors in Computer Systems; April 24-28, 1994; Boston, MA p. 413-414. [doi: [10.1145/259963.260531](https://doi.org/10.1145/259963.260531)]
 41. Love PED, Li H. Overcoming the problems associated with quality certification. *Constr Manag Econ* 2000 Mar;18(2):139-149. [doi: [10.1080/014461900370771](https://doi.org/10.1080/014461900370771)]
 42. Wohlin C. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. 2014 Presented at: 18th International Conference on Evaluation and Assessment in Software Engineering; May 13-14, 2014; London, UK p. 1-10. [doi: [10.1145/2601248.2601268](https://doi.org/10.1145/2601248.2601268)]

Abbreviations

- ACM:** Association for Computing Machinery
ASQ: After-Scenario Questionnaire
CENS: National Center for Health Information Systems
CORFO: Corporación de Fomento de la Producción
EHR: electronic health record
FONDECYT: Fondo Nacional de Desarrollo Científico y Tecnológico
FONDEF: Fondo de Fomento al Desarrollo Científico y Tecnológico
HCD: human-centered design
Health-ITUES: Health Information Technology Usability Evaluation Scale

HIS: health information system

IDeA: Investigación y Desarrollo en Acción

IEC: International Electrotechnical Commission

IEEE: Institute of Electrical and Electronics Engineers

ISO: International Organization for Standardization

MARS: Mobile App Rating Scale

mHealth: mobile health

NISTIR: National Institute of Standards and Technology Interagency/Internal Report

PICO: population, intervention, comparison, and output

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

SLR: systematic literature review

SUS: System Usability Scale

uMARS: end user version of the Mobile App Rating Scale

Edited by R Kukařka, A Mavragani; submitted 17.12.20; peer-reviewed by I Mircheva, S Nagavally, AUR Bacha, A Mac Cawley; comments to author 03.05.21; revised version received 30.07.21; accepted 04.12.21; published 08.03.22.

Please cite as:

Noël R, Taramasco C, Márquez G

Standards, Processes, and Tools Used to Evaluate the Quality of Health Information Systems: Systematic Literature Review

J Med Internet Res 2022;24(3):e26577

URL: <https://www.jmir.org/2022/3/e26577>

doi: [10.2196/26577](https://doi.org/10.2196/26577)

PMID: [35258469](https://pubmed.ncbi.nlm.nih.gov/35258469/)

©René Noël, Carla Taramasco, Gastón Márquez. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 08.03.2022. 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.

Review

Community Health Programs Delivered Through Information and Communications Technology in High-Income Countries: Scoping Review

Hannah Beks¹, MPH; Olivia King², PhD; Renee Clapham^{3,4}, PhD; Laura Alston^{1,5,6}, PhD; Kristen Glenister^{7,8}, PhD; Carol McKinstry⁹, PhD; Claire Quilliam⁷, PhD; Ian Wellwood¹⁰, PhD; Catherine Williams², GradCertClinEpi; Anna Wong Shee^{1,4}, PhD

¹School of Medicine, Deakin University, Geelong, Australia

²Barwon Health, Geelong, Australia

³St Vincents Health Australia, Melbourne, Australia

⁴Ballarat Health Services, Ballarat, Australia

⁵Colac Area Health, Colac, Australia

⁶Global Obesity Centre, Institute for Health Transformation, Deakin University, Geelong, Australia

⁷Department of Rural Health, University of Melbourne, Wangaratta, Australia

⁸Department of Rural Health, University of Melbourne, Shepparton, Australia

⁹La Trobe Rural Health School, La Trobe University, Bendigo, Australia

¹⁰Faculty of Health Sciences, Australian Catholic University, Ballarat, Australia

Corresponding Author:

Hannah Beks, MPH

School of Medicine

Deakin University

1 Gheringhap St

Geelong, 3220

Australia

Phone: 61 355633039

Email: hannah.beks@deakin.edu.au

Abstract

Background: The COVID-19 pandemic has required widespread and rapid adoption of information and communications technology (ICT) platforms by health professionals. Transitioning health programs from face-to-face to remote delivery using ICT platforms has introduced new challenges.

Objective: The objective of this review is to scope for ICT-delivered health programs implemented within the community health setting in high-income countries and rapidly disseminate findings to health professionals.

Methods: The Joanna Briggs Institute's scoping review methodology guided the review of the literature.

Results: The search retrieved 7110 unique citations. Each title and abstract was screened by at least two reviewers, resulting in 399 citations for full-text review. Of these 399 citations, 72 (18%) were included. An additional 27 citations were identified through reviewing the reference lists of the included studies, resulting in 99 citations. Citations examined 83 ICT-delivered programs from 19 high-income countries. Variations in program design, ICT platforms, research design, and outcomes were evident.

Conclusions: Included programs and research were heterogeneous, addressing prevalent chronic diseases. Evidence was retrieved for the effectiveness of nurse and allied health ICT-delivered programs. Findings indicated that outcomes for participants receiving ICT-delivered programs, when compared with participants receiving in-person programs, were either equivalent or better. Gaps included a paucity of co-designed programs, qualitative research around group programs, programs for patients and carers, and evaluation of cost-effectiveness. During COVID-19 and beyond, health professionals in the community health setting are encouraged to build on existing knowledge and address evidence gaps by developing and evaluating innovative ICT-delivered programs in collaboration with consumers and carers.

KEYWORDS

telemedicine; delivery of health care; pandemics; community health services; information and communications technology; mobile phone

Introduction

Background

Health professionals, working across community and acute health care settings, have responded rapidly to the COVID-19 pandemic by adopting information and communications technology (ICT) to continue delivering health programs [1-3]. Internationally, there has been an upward surge in the use of ICT to facilitate videoconferencing and telephone consultations to meet physical distancing requirements [4-6]. In Australia, this shift to telehealth in the community health setting required a temporary restructure to government funding models [7]. COVID-19 has been a catalyst for global adoption and focus on the prioritization of ICT in health, particularly in the community health setting (primary care, ambulatory care, home-based care, and outpatient hospital care) where primary and secondary prevention health programs are delivered [3,8-13].

Digital health, eHealth, and telehealth (including telemedicine) are terms used interchangeably and broadly defined as the use of ICT platforms for the remote delivery of health care to consumers [3,14,15]. Examples include videoconferencing and telephone consultations, web-based platforms, electronic health records, SMS text messaging, and smartphone apps (or mobile health, which can include telemonitoring platforms) [14]. Globally, there is increasing support for the use of ICT platforms to improve the accessibility of health services, particularly for health promotion and disease prevention [14,16]. This is evidenced by a surge in research evaluating the usability and effectiveness of ICT-delivered health [17,18], including programs addressing chronic disease risk factors [19-23], patient education and health literacy [24,25], and chronic disease self-management [18,26-28].

Barriers to the adoption of ICT platforms by health professionals are well documented and include a lack of ICT familiarity, lack of time to implement ICT programs, design and technical concerns, and attitudes toward ICT [29-32]. There has been little scope to address these barriers during the pandemic, where there has been a greater focus on the use of ICT in COVID-19 surveillance [33-35], and delivery of telehealth consultations [3,36]. To support health professionals in transitioning community health programs to remote delivery using ICT during COVID-19, a collaborative group was established between 4 Australian universities and 2 regional health services in April 2020. A review working group was formed, with the purpose of engaging directly with health professionals to understand knowledge gaps regarding program delivery using ICT. During the consultation phase (May to June 2020), health professionals voiced concerns regarding the transition of community health programs (particularly group programs) to an ICT platform and the potential for reduced program effectiveness. Similar

concerns have been shared by other health professionals internationally [37].

Approaches to undertaking reviews to inform evidence-based decision-making in health care vary [38]. Engaging stakeholders in the review process is suggested to generate more relevant review findings and enable prompt dissemination into practice [39]. An initial search was undertaken of MEDLINE Ovid, Cochrane Database of Systematic Reviews, Joanna Briggs Institute's (JBI) Evidence Synthesis, and PROSPERO for existing reviews (or proposed reviews) examining ICT-delivered health programs implemented in the community health care setting in high-income countries (HIC). No recent reviews were located that mapped the evidence for community health ICT-delivered programs, justifying the need for a scoping review [40]. The review was limited to HIC because advanced use of ICT platforms is more likely with similarities in resourcing [14]. Capturing a broad range of ICT platforms across various health disciplines and specialties was important for participating health professionals seeking to innovate and engage consumers in programs. Responding to these needs, researchers and health professionals in the review working group collaborated to develop the review question, objectives, and inclusion and exclusion criteria.

Review Questions and Objectives

The review question is as follows:

What is the evidence for the development and implementation of health programs delivered through ICT for consumers in the community health care settings in HIC?

The specific review objectives include the following:

1. to scope for evidence examining the development and implementation of ICT-delivered health programs in the community health care setting in HIC,
2. to scope for consumer co-design processes used to develop health programs,
3. to examine strategies to facilitate the sharing of consumer lived experience and peer interaction through an ICT platform, and
4. to scope for any andragogical or pedagogical principles or theories, informing program design.

Methods

Overview

This scoping review examined the evidence around ICT-delivered health programs implemented in HIC community health care settings. This review used the JBI's scoping review methodology [41]. Search terms were developed for the population, concept, and context. The review question, objectives, inclusion and exclusion criteria, and search strategies were developed and documented in advance (Section S1 in

[Multimedia Appendix 1](#) [41-141]). The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) was adhered to (Table S1 in [Multimedia Appendix 1](#)) [42-141].

Search Strategy

The JBI 3-step search process was used [142]. A preliminary search was undertaken in Ovid MEDLINE and CINAHL using keywords. A tailored search was then developed for each information source using keywords. For databases, a combination of Boolean operators, truncations, and Medical Subject Headings were used to form search strings ([Multimedia Appendix 1](#)). Health librarian assistance was obtained for developing the initial Ovid MEDLINE search strategy and translating searches. Reference lists of included studies were also searched for additional studies.

The databases searched included Ovid MEDLINE, CINAHL (EBSCOhost), Embase (Elsevier), and Cochrane Database of Systematic Reviews (Table S2 in [Multimedia Appendix 1](#)).

Table 1. Inclusion and exclusion criteria.

	Inclusion criteria	Exclusion criteria
Population	Health programs delivered for infants, children, young people, and adults, including those delivered for consumers, carers, and family or friends of consumers	No exclusions
Concept	Health programs (interventions, models of health care, and services, including, but not limited to, health education, self-management, health promotion and rehabilitation for secondary prevention of disease) delivered by health professionals (including psychologists, speech therapists, speech pathologists, occupational therapists, physiotherapists, physical therapists, podiatrists, exercise physiologists, dietitians, social workers, audiologists, nurses, and doctors) addressing health conditions including, but not limited to, chronic disease (eg, cardiovascular disease, respiratory disease, diabetes, renal disease, cancer, and mental illness) or risk factors for developing chronic disease including, but not limited to, obesity, physical inactivity, poor health literacy, and alcohol misuse using information and communications technology (eg, mobile health, eHealth, telehealth, web-based interventions, and digital health)	Infectious disease screening and surveillance programs, antenatal and postnatal programs, with the exception of gestational diabetes mellitus and postoperative rehabilitation programs
Context	Health programs implemented in the community health context in high-income countries (according to the Organization for Economic Co-operation and Development criteria), including primary care clinics and hospital outpatient clinics	Programs delivered in low- and middle-income countries

Study Selection and Data Extraction

Searches were undertaken with the assistance of librarians skilled in systematic reviews. Citations were imported into Covidence (Veritas Health Innovation) for screening. Titles and abstracts were screened independently by at least two reviewers with conflicts resolved through mediation with an independent reviewer. All authors were involved in either screening, resolving conflicts, or both. Authors only resolved conflicts for citations that they did not screen. Full-text review and data extraction was then undertaken. For articles not meeting the inclusion criteria, reasons were noted (Table S4 in [Multimedia Appendix 1](#)). Reference lists of the included citations were screened for additional literature. Data extraction was tabulated (Section S1 in [Multimedia Appendix 1](#)), and findings were

synthesized using a descriptive approach informed by the review objectives [41]. Consistent with scoping review methods and to enable rapid dissemination of findings, a quality assessment of the studies was not undertaken [143,144].

Inclusion Criteria and Exclusion Criteria

The literature was selected according to the inclusion and exclusion criteria presented in [Table 1](#)). Health programs (excluding infectious disease screening, surveillance, antenatal and postnatal, and postoperative rehabilitation programs) delivered by a health professional using an ICT platform to all populations (including carers and family members) in the community health context of HIC, as defined by the Organization for Economic Co-operation and Development (OECD) [142], were included. All types of literature published from January 1, 2010, to June 16, 2020, were included to capture a broad range of ICT platforms and health programs. Only studies published in English were included because of resource constraints.

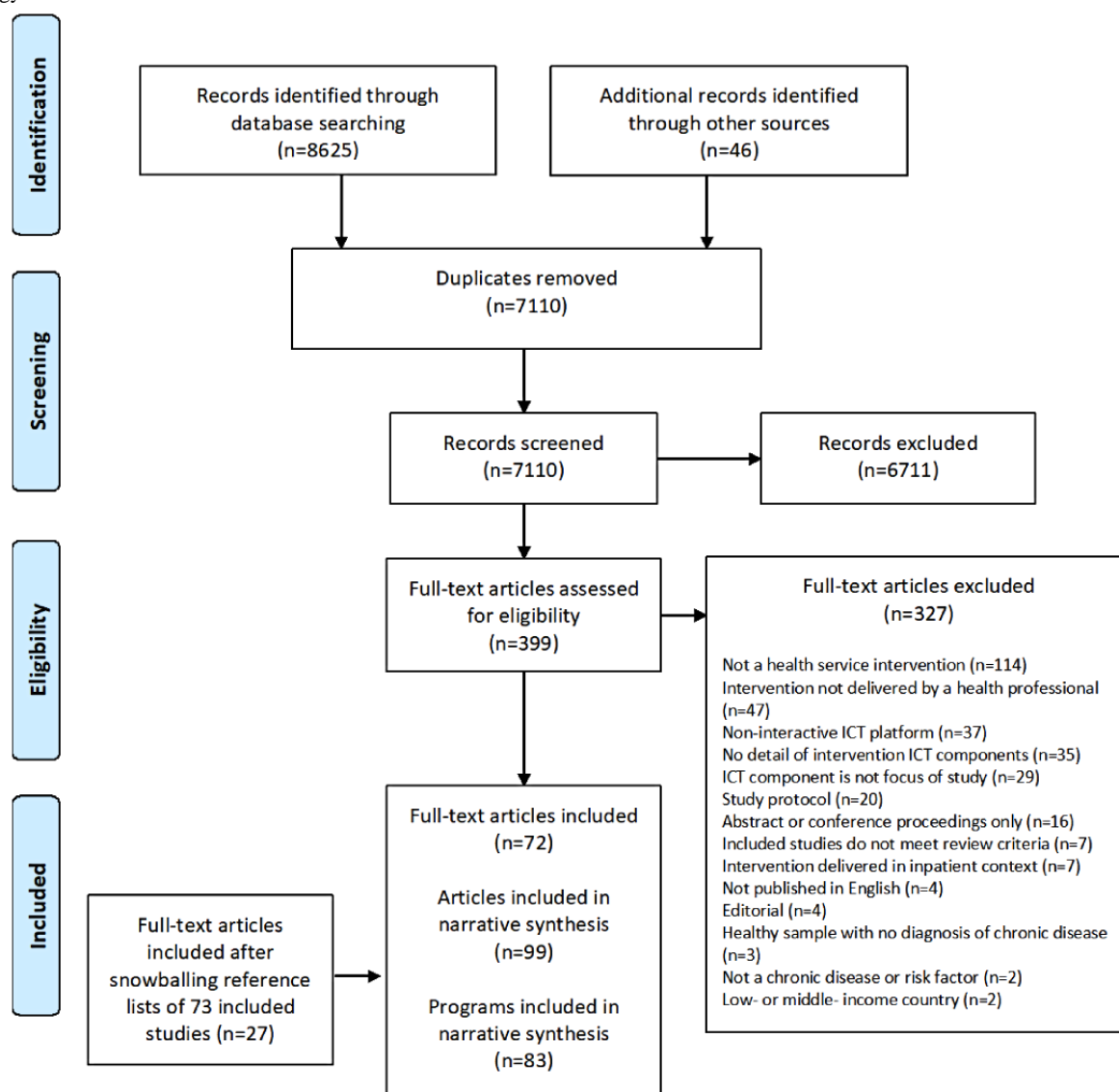
synthesized using a descriptive approach informed by the review objectives [41]. Consistent with scoping review methods and to enable rapid dissemination of findings, a quality assessment of the studies was not undertaken [143,144].

Results

Overview

Of the 399 citations eligible for full-text screening, 72 (18%) met the inclusion criteria. An additional 27 citations were identified from the reference lists of the included citations, resulting in 99 citations examining 83 programs ([Figure 1](#)). Reasons for exclusion were provided (Table S4 in [Multimedia Appendix 1](#)).

Figure 1. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flow diagram. ICT: information and communications technology.



Heterogeneity of Programs Using ICT Platforms

The included health programs (n=83) were heterogeneous in design and use of ICT platforms, addressing a variety of chronic diseases (cancer, 3/83, 4%; cardiovascular disease [CVD], 12/83, 14%; diabetes [including gestational diabetes], 30/83, 36%; chronic obstructive pulmonary disease [COPD], 14/83, 17%; other chronic diseases, 11/83, 13%; and chronic pain, 2/83, 2%) and risk factors for developing chronic disease (11/83, 13%; Table S5 in [Multimedia Appendix 1](#)). The most frequently used ICT platform for program delivery was the telephone (24/83, 29%) and then internet-based platforms (21/83, 25%), telehealth (telemonitoring; 15/83, 18%), and videoconferencing (11/83, 13%). Some programs used a combination of ICT: telephone and internet-based platforms (1/83, 1%); telephone and mobile apps (2/83, 2%); telemonitoring and an internet-based platform (6/83, 7%); and telehealth (telemonitoring), videoconferencing, and telephone (2/83, 2%). Most programs were delivered by nurses (30/83, 36%) or a multidisciplinary health care team (24/83, 29%), dietitians (8/83, 10%), physiotherapists (7/83, 8%), diabetes educators (4/83, 5%), and psychologists (4/83,

5%). Diverse community health settings were captured where the programs were delivered. Most programs were delivered in outpatient hospital settings (51/83, 61%), followed by home-based settings (12/83, 15%; delivered by other community health organizations that were not primary care practices or hospitals), primary care practices (10/83, 12%), and other community health centers, including multidisciplinary centers (7/83, 8%) and community cancer centers (3/83, 4%).

The included health programs were from 19 OECD HIC. The United States had the highest number of programs (31/83, 37%), followed by Australia (14/83, 17%), Canada (7/83, 8%), Spain (5/83, 6%), the United Kingdom (5/83, 6%), Denmark (4/83, 5%), Norway (3/83, 4%), Italy (3/83, 4%), the Netherlands (2/83, 2%), Belgium (2/83, 2%), Taiwan (2/83, 2%), Greece (2/83, 2%), France (2/83, 2%), Japan (1/83, 1%), Finland (1/83, 1%), Germany (1/83, 1%), South Korea (1/83, 1%), Singapore (1/83, 1%), and Switzerland (1/83, 1%). A total of 2 programs were implemented in >1 country, accounting for 88 sites of program implementation across all included studies [76,86].

Program Design: Group Programs, Co-design, and Guiding Theories

The programs primarily targeted only patients (76/83, 92%). Fewer programs were for patients and carers (7/83, 8%) and included 2 programs for cancer management [43,44], 1 telemonitoring program for CVD [56], 1 rehabilitation program for acquired brain injury [61], 1 pediatric asthma management program [64,65], 1 coping skills training program for COPD [78], and 1 self-management program for psychological distress [133].

Of the 83 programs, 16 (19%) were either delivered to groups of participants or included a component that involved groups of participants. Of these 16 programs, 5 (31%) targeted diabetes education, self-management, and behavior change coaching [101,102,108,116,125]; 4 (25%) programs were CVD rehabilitation (secondary prevention) or counseling programs [46,47,52,53,55,57]; 4 (25%) addressed risk factors for chronic disease through education and behavior change coaching [131,132,136,141]; 1 (6%) involved group cognitive behavioral therapy (CBT) for participants experiencing insomnia [66]; 1 (6%) involved pharmacist-led group education for hepatitis C [73]; and 1 (6%) involved group education for osteoarthritis [70].

No studies included strategies to facilitate the sharing of consumer lived experience and peer interaction in group ICT-delivered programs. A qualitative study evaluating 1 group program (CVD rehabilitation program) reported that participants engaged in group sessions but did not provide information regarding participants' experiences [46]. There was limited information of any co-design processes used with consumers or participants to develop programs. Only 2 studies investigating 2 different programs mentioned collaboration with consumers or community organizations to develop interventions; however, no detail about the collaboration was provided [70,119].

None of the studies used specific andragogical or pedagogical principles to inform the delivery of ICT programs to adult or child participants. A total of 12 citations referred explicitly to health behavior theories that informed program development or delivery. Constructs of social cognitive theory (SCT) were used to inform a diabetes self-management support program (Health Education Access Through Information Technology and Utilization Program) [123], a diabetes telemedicine program [111], a pedometer-based intervention for the secondary prevention of CVD [50], a telephone-based Living Well with Diabetes program [104], a telephone-based symptom management program for people with lung cancer and their carers [76], and a telephone-based health coaching program for the secondary prevention of CVD [58,59]. Strategies were implemented to optimize program participation and adherence by promoting SCT constructs (eg, self-efficacy). Examples of strategies included supporting participants to engage in goal setting (eg, related to physical activity) [50], encouraging participants to seek support and rewarding achievements [104], and equipping participants with skills (through cognitive restructuring, problem solving, or self-soothing) to enhance self-efficacy [76].

Other theories included self-determination theory, which informed the development of a telephone-based coaching program targeting physical activity and quality of life for inactive adults through self-management [132]. Using self-determination theory as a conceptual framework, the program integrated motivational interviewing and CBT approaches to coaching [132]. The chronic care model developed by Wagner et al [145] and the transtheoretical model [146] were also used to guide a diabetes self-management education program [119], enabling self-management education and management goals to be provided and set specifically for the stage of change participants were at. The transtheoretical model was also used to inform the content and delivery of pediatric asthma management programs delivered to children and their carers [64,65] and a telehealth diabetes self-management program, along with the health belief model [102].

Research Evidence: Study Designs, Findings, and Limitations

Heterogeneity was evident in the research design of included citations ($n=99$) when evaluating the effectiveness, feasibility, or acceptability of the included programs ($n=83$; Table S6 in [Multimedia Appendix 1](#)). Most studies used a randomized controlled trial (RCT) design (58/99, 59%), followed by a single cohort study design (12/99, 12%), a cohort study with 2 or more groups (7/99, 7%), a qualitative design (5/99, 5%), an economic evaluation of an RCT (4/99, 4%), a mixed methods study design (3/99, 3%), or a survey design (2/99, 2%). The remaining citations used other non-RCTs or experimental study designs (8/99, 8%).

Primary and secondary outcomes, and approaches to measuring outcomes (eg, use of validated questionnaires or devices) varied between studies and conditions (Table S6 in [Multimedia Appendix 1](#)). For RCTs, the reported effect was categorized as positive (ICT intervention was effective or more effective than control), neutral (effects were equivalent to control), or negative (ICT intervention was not effective or less effective than control) where appropriate, to provide an indication of the effectiveness of programs using ICT platforms. Of the 58 studies able to be categorized, 30 (52%) reported positive effects on the primary and secondary outcomes attributed to the ICT intervention, when compared with the control group, whereas 28 (48%) studies reported a neutral effect. No RCTs reported that outcomes were worse in the ICT intervention group than in the control group. Owing to the heterogeneity of primary and secondary outcome measures and program design, the most frequently reported outcome measures for condition groups used in RCTs are reported in [Table 2](#), with the effects categorized. From the studies included in this table, there was consistency in the findings of RCTs of COPD programs reporting on health service use outcome measures. The effect of programs on the rate of hospitalization of the ICT intervention and control groups were found to be neutral. However, for RCTs of programs using clinical, anthropometric, or physical activity outcome measures, there was a mix of positive and neutral effects. The length of the final follow-up periods in RCTs ranged from 6 weeks to 5 years (with a median follow-up period of 12, IQR 6-15 months).

Table 2. Most frequently reported primary outcome measures in included RCTs^a.

Study	Reported effect and results
Outcome measure: HbA_{1c}^b (diabetes programs)	
Baron et al [94]	Neutral: Program did not achieve a clinically significant reduction in HbA _{1c} .
Blackberry et al [96]	Neutral: At 18-months follow-up, the effect on HbA _{1c} did not differ between the intervention and control (mean difference 0.2, 95%CI -0.2 to 0.2; <i>P</i> =.84).
Buysse et al [97]	Positive: Both groups received tele-education at different time points (delayed access [control] and immediate access [study group]) and demonstrated an overall significant impact of tele-education on HbA _{1c} reduction (-0.5% control and -0.4% study group, respectively).
Carter et al [98]	Positive: Patients enrolled in intervention were 4.58 times more likely to achieve an HbA _{1c} target <7%.
Charpentier et al [99]	Positive: At 6 months, mean HbA _{1c} was lower in the intervention group than in the control group (8.41 vs 9.10, respectively).
Davis et al [102]	Positive: A significant reduction in HbA _{1c} was found in the intervention group, compared with usual care (9.4 to 8.2 in the intervention group, compared with 8.8 to 8.6 in usual care).
Fountoulakis et al [107]	Positive: Significant reduction in HbA _{1c} in the intervention group at 3 and 6 months, when compared with that in the control group.
Greenwood et al [108]	Positive: The intervention group had a statistically significant difference of 0.41 percentage points at 6 months when compared with the control group.
Klingeman et al [117]	Positive: Average HbA _{1c} reduced by 1.7% in the intervention group, compared with 0.3% in the control group.
Sood et al [124]	Neutral: No statistically significant differences between the intervention and control groups at 18 months.
Varney et al [127]	Positive: The intervention group experienced a greater mean change in adjusted HbA _{1c} than the controls between baseline and 12 months; however, this was not sustained.
Wakefield et al [129]	Neutral: Participants in the intervention group experienced decreased HbA _{1c} during the 6-month intervention period when compared with the control group; however, 6 months after the intervention was withdrawn, the intervention groups were comparable with the control group.
Weinstock et al [113]	Positive: Intervention was associated with improved HbA _{1c} over 5 years, when compared with control.
Wild et al [110]	Positive: Clinically and statistically significant improvements were observed in the intervention group at 9 months, when compared with the control group.
Outcome measure: rate of hospitalization (COPD^c programs)	
Antoniades et al [75]	Neutral: No significant difference between the intervention and control groups at 12 months.
Blumenthal et al [78]	Neutral: No significant difference between the intervention and control groups up to 4.4 years follow-up.
Fairbrother et al [84]	Neutral: No significant difference between the intervention and control groups at 12 months.
Pinnock et al [85]	Neutral: No significant difference between the intervention and control groups at 12 months.
Kessler et al [86]	Neutral: No significant difference between the intervention and control groups at 12 months.
Tabak et al [89]	Neutral: No significant difference between the intervention and control groups at 2 months.
Outcome measure: PA^d or capacity (cardiovascular disease programs)	
Lear et al [47]	Positive: Intervention group participants who received support from a health professional through an internet-based platform had a greater increase in maximal time on the treadmill by 45.7 seconds (95% CI 1.04-90.48) compared with the usual care group over the 16 months (<i>P</i> =.045).
Furber et al [50]	Positive: After the 6-week intervention, improvements in total PA time, total PA sessions, walking time, and walking sessions were all significantly greater in the intervention group who received telephone support than in the control who received 2 education pamphlets and no support via telephone.
Hawkes et al [59]	Neutral: No significant difference between the PA of participants in the intervention and control groups at 6 months follow-up.
Hwang et al [52,53]	Neutral: No difference was found between the PA of participants receiving the telerehabilitation intervention when compared with the control group who received center-based care, and it was less costly than center-based heart failure rehabilitation.
Nolan et al [57]	Positive: More telehealth participants than control participants reported adherence to exercise and diet after treatment at a 6-month follow-up.

Study	Reported effect and results
Outcome measure: weight loss or prevention of weight gained (risk factors for chronic disease programs)	
Ferrara et al [135]	Positive: Compared with those receiving usual care, women in the lifestyle intervention had reduced weekly rate of gestational weight gain (mean 0.26 vs 0.32 kg/week).
Padwal et al [138]	Neutral: Face to-face or web-based delivery of intensive self-management program was no more effective than the once off provision of educational materials and were more costly.
Weinstock et al [141]	Positive: Mean percent weight loss at 2-year follow-up was higher for the conference call group than for the individual call group (–5.6% compared with –1.8%).

^aRCT: randomized controlled trial.

^bHbA_{1c}: glycated hemoglobin A_{1c}.

^cCOPD: chronic obstructive pulmonary disease.

^dPA: physical activity.

Of the 7 studies using qualitative inquiry (including mixed methods studies using qualitative inquiry), 3 (43%) studies examined the attitudes of participants (a videoconferencing education workshop for inflammatory arthritis, a COPD telemonitoring program, and a telemonitoring program for diabetes) [71,88,109], 2 (29%) examined perceptions of a T2DM smartphone app [118,121], 1 (14%) measured the patient experience of being involved in a web-based cardiac rehabilitation program [46], and 1 (14%) examined the perceptions of both patients and health professionals involved with a COPD telemonitoring service [83]. Themes varied but generally related to the accessibility of ICT programs [46] and general participant satisfaction [88]. A study also reported no difference in feedback obtained from participants who attended an in-person program compared with those who attended videoconferencing [71]. Another study reported limitations of using ICT platforms, including frustration with using smartphones [118], whereas other studies reported that technology was acceptable [83,88,109,121].

Studies providing an economic evaluation of an ICT-delivered program, in conjunction with either an RCT [49,52,54,132] or a case-control study [51], supported the potential for the cost-effectiveness of ICT-delivered programs when compared with in-person programs. When examining telerehabilitation for CVD, Hwang et al [52] found the intervention to be as effective and less costly than center-based rehabilitation. Ho et al [51] reported that a telehealth program for CVD was more cost-effective and more likely to prevent hospitalizations than usual care. However, a telemonitoring program for CVD was reportedly not cost-effective because the intervention had higher costs (including equipment costs) than usual care, and no significant difference was found in quality-adjusted life years [54].

Research limitations frequently reported included high attrition rates, small sample sizes (or not statistically powered for outcome measures), and limited external validity. The total attrition rates of RCTs ranged from 1% to 63%, with a median attrition rate of 18% (IQR 10%-25%). Difficulty in recruiting participants was also reported by some researchers. An RCT conducted a survey of why participants declined to participate in the trial and found personal reasons and concerns with technology were frequently cited by respondents [54].

Discussion

Principal Findings

This review provides a broad overview of research examining ICT-delivered programs implemented in the community health setting in 19 countries, providing a sample of programs from 24% (19/80) of OECD HIC [144]. The highest proportion of included ICT-delivered programs was implemented in the United States, the country with the highest financial investment in health care (16.9% of gross domestic product in 2018) [147] and a growing investment in digital health [148]. Although this review was limited to programs implemented by OECD HIC, other studies have identified a surge in ICT programs and innovations in low- and middle-income countries [149-151]. Because of the COVID-19 pandemic, it is anticipated that ICT-delivered health programs and innovations will continue to increase as global health care systems are transformed [152].

Included programs and citations were diverse, addressing a range of chronic diseases and risk factors, using a variety of ICT platforms delivered by different health professionals across different community health settings. Programs mostly targeted highly prevalent chronic diseases and risk factors, such as CVD, COPD, diabetes, and obesity or being overweight [153,154], and were delivered in the outpatient hospital setting. The need to facilitate a greater adoption of ICT in other community health settings (eg, primary care practices) has been identified by other international research and is supported by the review findings [155]. Furthermore, there were few self-management and education programs addressing cancer and mental health conditions, other chronic diseases that pose a significant burden on global populations [153]. During the COVID-19 pandemic, the need for improved accessibility to mental health programs has also been identified [156]. There were also few programs implemented for patients and carers. There is an increasing focus on the importance of carer engagement, particularly for dementia care [157] and mental health [158], and research around the role of ICT programs in supporting carers [159].

A high proportion of programs were delivered using the telephone, internet, and telemonitoring. With a surge in the use of mobile health technologies through smartphone apps and other innovations (eg, activity monitoring devices), this finding suggests that the telephone remains an important ICT platform for improving patient accessibility to health professionals,

particularly for self-management and behavior change coaching. This is evident by the use of telehealth during COVID-19 in countries such as Australia, where telephone consultations have had a higher uptake in primary care settings compared with videoconferencing delivered via web-based platforms [160]. The usefulness of videoconferencing for delivering group education, behavior change coaching, and self-management programs is also indicated by the review findings. Although this review reports little about the acceptability of ICT-delivered group programs and strategies to facilitate peer interaction, other reviews have found that group programs delivered through videoconferencing have been acceptable and feasible to participants [161]. Future research needs to examine how to facilitate group interaction in ICT programs [162].

Although the included studies had a range of research designs (a finding of another systematic review examining emergent eHealth interventions [163]), the findings from this review supported the effectiveness of nurse-led ICT programs in improving pain associated with cancer [43], improving quality-of-life outcomes and reducing hospital admissions for patients with CVD [56], improving health outcomes for patients with CVD [58,59], and improving quality of life in carers of children with asthma [64,65]. Findings also indicated the effectiveness of ICT-delivered programs by allied health professionals, including a telenutrition program delivered by dietitians [60] and a chronic pain program delivered by physiotherapists [91]. The results from included RCTs comparing participant outcomes of an ICT program to a control group (receiving mostly in-person care) were either equivalent or better for ICT programs. Other reviews examining ICT interventions, such as nurse and allied health, delivered telehealth interventions [164], and electronic CBT [165], also concluded that delivering health interventions through ICT platforms does not lead to poorer health outcomes for patients.

Substantial gaps in research evidence relating to ICT programs delivered in the community health setting by health professionals were identified. There were few co-designed ICT programs (and no documentation of co-design processes) and no reference to specific pedagogical or andragogical educational principles guiding program delivery—gaps identified by other reviews [166,167]. Engaging stakeholders in program development through co-design processes is thought to create programs that are more useful and acceptable to end users [168]. Some programs were developed or guided by theories; however, further research is required to examine whether using theories (eg, SCT) to develop and guide programs results in better

outcomes for participants [169]. Few studies have examined participant acceptability, experience, and perceptions of ICT programs through qualitative inquiry. However, qualitative findings resonate with other reviews that have found that participants are generally satisfied with telehealth [170]. Findings indicate that there is a need for greater consumer engagement in the process of developing ICT programs and evaluating effectiveness [171]. There is also a need for more economic evaluations of ICT programs delivered in the community health setting, which is also lacking in broader health services research [172,173].

Engaging with health professionals to understand knowledge gaps regarding community health ICT program delivery during COVID-19 and codevelopment of the scoping review question, objectives, and inclusion and exclusion criteria are strengths of this review. A summary of review findings was rapidly disseminated to health professionals involved, and findings were discussed during a short webinar. The limitations of the review include only a brief search of international gray literature due to the need to rapidly disseminate findings to health professionals. Undertaking a more thorough search of the international gray literature could have minimized publication bias. There is potential that relevant citations were not included in the review owing to this constraint. Despite this, every effort was made to review the reference lists of included citations for additional studies. Studies published in a language other than English were not captured by this review owing to resource constraints.

Conclusions

This review identified heterogeneity in available evidence examining ICT-delivered programs in community health settings in HIC. There is promising evidence for the effectiveness of nurse and allied health delivered ICT programs. From RCTs, outcomes for participants receiving ICT programs, compared with those receiving in-person programs, were either equivalent or better. Gaps identified included a paucity of co-designed programs; qualitative research relating to consumer acceptability, experience, and interactions in group programs; and cost-effectiveness of ICT programs and programs targeting patients and carers. It is expected that because of COVID-19, there will be a surge in the innovation, development, and evaluation of community health programs delivered using ICT platforms, providing an opportunity for health professionals and researchers to build on existing knowledge and address evidence gaps.

Acknowledgments

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

We acknowledge health professionals who participated in the review working group and other collaborators: Luke Evans, Rachel Jamieson, Jessica McDonald, Sarah McKinnon, Fiona Orton, Michael Pang, Rishni Perera, Helen Ramsay, Annabel Brennan, Jeremy Caunt, Marg Dawson, Jaquie Evans-Barker, Tahlia Farragher, Valetta Fraser, Karen Goodwin, Kendl Gorrie, Sarah Kelley, Georgie Kemp, Rachel Maher, Kylie McKenzie, Ellie Orr, Michele Pearson, Andrea Pinch, Peter Roche, Amy Smith, Michelle Stewart, Acushla Thompson, Sue Titcumb, Angela Tudor, and Renee van der Horst. We also acknowledge Deakin University Librarians Fiona Russell, Chrissy Freestone, and Jill Stephens, whose assistance was invaluable in the development of the database search strategies for this review.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Included studies characteristics, search strategies, protocol, excluded studies, and PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.

[[DOCX File, 180 KB - jmir_v24i3e26515_app1.docx](#)]

References

1. Bielicki JA, Duval X, Gobat N, Goossens H, Koopmans M, Tacconelli E, et al. Monitoring approaches for health-care workers during the COVID-19 pandemic. *Lancet Infect Dis* 2020;20(10):e261-e267 [FREE Full text] [doi: [10.1016/S1473-3099\(20\)30458-8](https://doi.org/10.1016/S1473-3099(20)30458-8)] [Medline: [32711692](https://pubmed.ncbi.nlm.nih.gov/32711692/)]
2. Kishore S, Hayden M. Community health centers and Covid-19 - time for congress to act. *N Engl J Med* 2020;383(8):e54. [doi: [10.1056/NEJMp2020576](https://doi.org/10.1056/NEJMp2020576)] [Medline: [32589372](https://pubmed.ncbi.nlm.nih.gov/32589372/)]
3. Fisk M, Livingstone A, Pit SW. Telehealth in the context of COVID-19: changing perspectives in Australia, the United Kingdom, and the United States. *J Med Internet Res* 2020;22(6):e19264 [FREE Full text] [doi: [10.2196/19264](https://doi.org/10.2196/19264)] [Medline: [32463377](https://pubmed.ncbi.nlm.nih.gov/32463377/)]
4. Webster P. Virtual health care in the era of COVID-19. *Lancet* 2020;395(10231):1180-1181 [FREE Full text] [doi: [10.1016/S0140-6736\(20\)30818-7](https://doi.org/10.1016/S0140-6736(20)30818-7)] [Medline: [32278374](https://pubmed.ncbi.nlm.nih.gov/32278374/)]
5. Thomas EE, Haydon HM, Mehrotra A, Caffery LJ, Snoswell CL, Banbury A, et al. Building on the momentum: sustaining telehealth beyond COVID-19. *J Telemed Telecare* (forthcoming) 2020:1357633X20960638. [doi: [10.1177/1357633X20960638](https://doi.org/10.1177/1357633X20960638)] [Medline: [32985380](https://pubmed.ncbi.nlm.nih.gov/32985380/)]
6. Koonin LM, Hoots B, Tsang CA, Leroy Z, Farris K, Jolly T, et al. Trends in the use of telehealth during the emergence of the COVID-19 pandemic - United States, January-March 2020. *MMWR Morb Mortal Wkly Rep* 2020;69(43):1595-1599 [FREE Full text] [doi: [10.15585/mmwr.mm6943a3](https://doi.org/10.15585/mmwr.mm6943a3)] [Medline: [33119561](https://pubmed.ncbi.nlm.nih.gov/33119561/)]
7. COVID-19 temporary MBS telehealth services. Australian Government Department of Health. 2020. URL: <http://www.mbsonline.gov.au/internet/mbsonline/publishing.nsf/Content/Factsheet-TempBB> [accessed 2020-11-04]
8. McMahon M, Nadigel J, Thompson E, Glazier RH. Informing Canada's health system response to COVID-19: priorities for health services and policy research. *Healthc Policy* 2020;16(1):112-124 [FREE Full text] [doi: [10.12927/hcpol.2020.26249](https://doi.org/10.12927/hcpol.2020.26249)] [Medline: [32813643](https://pubmed.ncbi.nlm.nih.gov/32813643/)]
9. Bloem BR, Dorsey ER, Okun MS. The coronavirus disease 2019 crisis as catalyst for telemedicine for chronic neurological disorders. *JAMA Neurol* 2020;77(8):927-928. [doi: [10.1001/jamaneurol.2020.1452](https://doi.org/10.1001/jamaneurol.2020.1452)] [Medline: [32329796](https://pubmed.ncbi.nlm.nih.gov/32329796/)]
10. Duckett S. What should primary care look like after the COVID-19 pandemic? *Aust J Prim Health* 2020;26(3):207-211. [doi: [10.1071/PY20095](https://doi.org/10.1071/PY20095)] [Medline: [32454003](https://pubmed.ncbi.nlm.nih.gov/32454003/)]
11. Blandford A, Wesson J, Amalberti R, AlHazme R, Allwihan R. Opportunities and challenges for telehealth within, and beyond, a pandemic. *Lancet Glob Health* 2020;8(11):e1364-e1365 [FREE Full text] [doi: [10.1016/S2214-109X\(20\)30362-4](https://doi.org/10.1016/S2214-109X(20)30362-4)] [Medline: [32791119](https://pubmed.ncbi.nlm.nih.gov/32791119/)]
12. Schwamm LH, Estrada J, Erskine A, Licurse A. Virtual care: new models of caring for our patients and workforce. *Lancet Digit Health* 2020;2(6):e282-e285 [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30104-7](https://doi.org/10.1016/S2589-7500(20)30104-7)] [Medline: [32382724](https://pubmed.ncbi.nlm.nih.gov/32382724/)]
13. Smith AC, Thomas E, Snoswell CL, Haydon H, Mehrotra A, Clemensen J, et al. Telehealth for global emergencies: implications for coronavirus disease 2019 (COVID-19). *J Telemed Telecare* 2020;26(5):309-313 [FREE Full text] [doi: [10.1177/1357633X20916567](https://doi.org/10.1177/1357633X20916567)] [Medline: [32196391](https://pubmed.ncbi.nlm.nih.gov/32196391/)]
14. Global diffusion of eHealth: making universal health coverage achievable. World Health Organization. 2016. URL: https://www.who.int/goe/publications/global_diffusion/en/ [accessed 2020-11-04]
15. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. *J Med Internet Res* 2005;7(1):e1 [FREE Full text] [doi: [10.2196/jmir.7.1.e1](https://doi.org/10.2196/jmir.7.1.e1)] [Medline: [15829471](https://pubmed.ncbi.nlm.nih.gov/15829471/)]
16. Digital health. Seventy-First World Health Assembly. 2018. URL: https://apps.who.int/gb/ebwha/pdf_files/WHA71/A71_R7-en.pdf?ua=1 [accessed 2020-11-04]
17. Inal Y, Wake JD, Guribye F, Nordgreen T. Usability evaluations of mobile mental health technologies: systematic review. *J Med Internet Res* 2020;22(1):e15337 [FREE Full text] [doi: [10.2196/15337](https://doi.org/10.2196/15337)] [Medline: [31904579](https://pubmed.ncbi.nlm.nih.gov/31904579/)]
18. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. *Annu Rev Public Health* 2015;36:393-415 [FREE Full text] [doi: [10.1146/annurev-publhealth-031914-122855](https://doi.org/10.1146/annurev-publhealth-031914-122855)] [Medline: [25785892](https://pubmed.ncbi.nlm.nih.gov/25785892/)]
19. Marin TS, Kourbelis C, Foote J, Newman P, Brown A, Daniel M, et al. Examining adherence to activity monitoring devices to improve physical activity in adults with cardiovascular disease: a systematic review. *Eur J Prev Cardiol* 2019;26(4):382-397. [doi: [10.1177/2047487318805580](https://doi.org/10.1177/2047487318805580)] [Medline: [30322268](https://pubmed.ncbi.nlm.nih.gov/30322268/)]
20. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. *J Cardiovasc Nurs* 2013;28(4):320-329 [FREE Full text] [doi: [10.1097/JCN.0b013e318250a3e7](https://doi.org/10.1097/JCN.0b013e318250a3e7)] [Medline: [22635061](https://pubmed.ncbi.nlm.nih.gov/22635061/)]

21. Su JJ, Yu DS, Paguio JT. Effect of eHealth cardiac rehabilitation on health outcomes of coronary heart disease patients: a systematic review and meta-analysis. *J Adv Nurs* 2020;76(3):754-772. [doi: [10.1111/jan.14272](https://doi.org/10.1111/jan.14272)] [Medline: [31769527](https://pubmed.ncbi.nlm.nih.gov/31769527/)]
22. 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](https://doi.org/10.1177/2055207619826685)] [Medline: [30783535](https://pubmed.ncbi.nlm.nih.gov/30783535/)]
23. Brørs G, Pettersen TR, Hansen TB, Fridlund B, Hølvold LB, Lund H, et al. Modes of e-Health delivery in secondary prevention programmes for patients with coronary artery disease: a systematic review. *BMC Health Serv Res* 2019;19(1):364 [FREE Full text] [doi: [10.1186/s12913-019-4106-1](https://doi.org/10.1186/s12913-019-4106-1)] [Medline: [31182100](https://pubmed.ncbi.nlm.nih.gov/31182100/)]
24. Halldorsdottir H, Thoroddsen A, Ingadottir B. Impact of technology-based patient education on modifiable cardiovascular risk factors of people with coronary heart disease: a systematic review. *Patient Educ Couns* 2020;103(10):2018-2028. [doi: [10.1016/j.pec.2020.05.027](https://doi.org/10.1016/j.pec.2020.05.027)] [Medline: [32595027](https://pubmed.ncbi.nlm.nih.gov/32595027/)]
25. Jacobs RJ, Lou JQ, Ownby RL, Caballero J. A systematic review of eHealth interventions to improve health literacy. *Health Informatics J* 2016;22(2):81-98 [FREE Full text] [doi: [10.1177/1460458214534092](https://doi.org/10.1177/1460458214534092)] [Medline: [24916567](https://pubmed.ncbi.nlm.nih.gov/24916567/)]
26. Choi W, Wang S, Lee Y, Oh H, Zheng Z. A systematic review of mobile health technologies to support self-management of concurrent diabetes and hypertension. *J Am Med Inform Assoc* 2020;27(6):939-945 [FREE Full text] [doi: [10.1093/jamia/ocaa029](https://doi.org/10.1093/jamia/ocaa029)] [Medline: [32357368](https://pubmed.ncbi.nlm.nih.gov/32357368/)]
27. Low JK, Manias E. Use of technology-based tools to support adolescents and young adults with chronic disease: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2019;7(7):e12042 [FREE Full text] [doi: [10.2196/12042](https://doi.org/10.2196/12042)] [Medline: [31322129](https://pubmed.ncbi.nlm.nih.gov/31322129/)]
28. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. *J Med Internet Res* 2018;20(5):e162 [FREE Full text] [doi: [10.2196/jmir.9751](https://doi.org/10.2196/jmir.9751)] [Medline: [29728346](https://pubmed.ncbi.nlm.nih.gov/29728346/)]
29. Gagnon MP, Desmartis M, Labrecque M, Car J, Pagliari C, Pluye P, et al. Systematic review of factors influencing the adoption of information and communication technologies by healthcare professionals. *J Med Syst* 2012;36(1):241-277 [FREE Full text] [doi: [10.1007/s10916-010-9473-4](https://doi.org/10.1007/s10916-010-9473-4)] [Medline: [20703721](https://pubmed.ncbi.nlm.nih.gov/20703721/)]
30. Koivunen M, Saranto K. Nursing professionals' experiences of the facilitators and barriers to the use of telehealth applications: a systematic review of qualitative studies. *Scand J Caring Sci* 2018;32(1):24-44. [doi: [10.1111/scs.12445](https://doi.org/10.1111/scs.12445)] [Medline: [28771752](https://pubmed.ncbi.nlm.nih.gov/28771752/)]
31. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. *J Telemed Telecare* 2018;24(1):4-12 [FREE Full text] [doi: [10.1177/1357633X16674087](https://doi.org/10.1177/1357633X16674087)] [Medline: [29320966](https://pubmed.ncbi.nlm.nih.gov/29320966/)]
32. 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;37(12):1967-1974. [doi: [10.1377/hlthaff.2018.05125](https://doi.org/10.1377/hlthaff.2018.05125)] [Medline: [30633683](https://pubmed.ncbi.nlm.nih.gov/30633683/)]
33. Digital tools for COVID-19 contact tracing. World Health Organisation. 2020. URL: <https://www.who.int/publications/item/WHO-2019-nCoV-Contact-Tracing-Tools-Annex-2020.1> [accessed 2020-11-04]
34. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. *Nat Med* 2020;26(8):1183-1192. [doi: [10.1038/s41591-020-1011-4](https://doi.org/10.1038/s41591-020-1011-4)] [Medline: [32770165](https://pubmed.ncbi.nlm.nih.gov/32770165/)]
35. Whitelaw S, Mamas MA, Topol E, Van Spall HG. Applications of digital technology in COVID-19 pandemic planning and response. *Lancet Digit Health* 2020;2(8):e435-e440 [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30142-4](https://doi.org/10.1016/S2589-7500(20)30142-4)] [Medline: [32835201](https://pubmed.ncbi.nlm.nih.gov/32835201/)]
36. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. *BMC Public Health* 2020;20(1):1193 [FREE Full text] [doi: [10.1186/s12889-020-09301-4](https://doi.org/10.1186/s12889-020-09301-4)] [Medline: [32738884](https://pubmed.ncbi.nlm.nih.gov/32738884/)]
37. Jiménez-Rodríguez S, Santillán García A, Montoro Robles J, Rodríguez Salvador MD, Muñoz Ronda FJ, Arrogante O. Increase in video consultations during the COVID-19 pandemic: healthcare professionals' perceptions about their implementation and adequate management. *Int J Environ Res Public Health* 2020;17(14):5112 [FREE Full text] [doi: [10.3390/ijerph17145112](https://doi.org/10.3390/ijerph17145112)] [Medline: [32679848](https://pubmed.ncbi.nlm.nih.gov/32679848/)]
38. Polisena J, Garritty C, Kamel C, Stevens A, Abou-Setta AM. Rapid review programs to support health care and policy decision making: a descriptive analysis of processes and methods. *Syst Rev* 2015;4:26 [FREE Full text] [doi: [10.1186/s13643-015-0022-6](https://doi.org/10.1186/s13643-015-0022-6)] [Medline: [25874967](https://pubmed.ncbi.nlm.nih.gov/25874967/)]
39. Feldmann J, Puhan MA, Mütsch M. Characteristics of stakeholder involvement in systematic and rapid reviews: a methodological review in the area of health services research. *BMJ Open* 2019;9(8):e024587 [FREE Full text] [doi: [10.1136/bmjopen-2018-024587](https://doi.org/10.1136/bmjopen-2018-024587)] [Medline: [31420378](https://pubmed.ncbi.nlm.nih.gov/31420378/)]
40. Peters MD, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. *JBIEvid Synth* 2020;18(10):2119-2126. [doi: [10.11124/JBIES-20-00167](https://doi.org/10.11124/JBIES-20-00167)] [Medline: [33038124](https://pubmed.ncbi.nlm.nih.gov/33038124/)]
41. Peters MG, Godfrey C, McInerney P, Munn Z, Tricco AC, Khalil H. Chapter 11: Scoping reviews (2020 version). In: Aromataris E, Munn Z, editors. *JBIE manual for evidence synthesis*. Adelaide, Australia: Joanna Briggs Institute; 2020.
42. 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;169(7):467-473 [FREE Full text] [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]

43. Badger TA, Segrin C, Sikorskii A, Pasvogel A, Weihs K, Lopez AM, et al. Randomized controlled trial of supportive care interventions to manage psychological distress and symptoms in Latinas with breast cancer and their informal caregivers. *Psychol Health* 2020;35(1):87-106. [doi: [10.1080/08870446.2019.1626395](https://doi.org/10.1080/08870446.2019.1626395)] [Medline: [31189338](https://pubmed.ncbi.nlm.nih.gov/31189338/)]
44. Mosher CE, Winger JG, Hanna N, Jalal SI, Einhorn LH, Birdas TJ, et al. Randomized pilot trial of a telephone symptom management intervention for symptomatic lung cancer patients and their family caregivers. *J Pain Symptom Manage* 2016;52(4):469-482 [FREE Full text] [doi: [10.1016/j.jpainsymman.2016.04.006](https://doi.org/10.1016/j.jpainsymman.2016.04.006)] [Medline: [27401514](https://pubmed.ncbi.nlm.nih.gov/27401514/)]
45. Thomas ML, Elliott JE, Rao SM, Fahey KF, Paul SM, Miaskowski C. A randomized, clinical trial of education or motivational-interviewing-based coaching compared to usual care to improve cancer pain management. *Oncol Nurs Forum* 2012;39(1):39-49 [FREE Full text] [doi: [10.1188/12.ONF.39-49](https://doi.org/10.1188/12.ONF.39-49)] [Medline: [22201654](https://pubmed.ncbi.nlm.nih.gov/22201654/)]
46. Banner D, Lear S, Kandola D, Singer J, Horvat D, Bates J, et al. The experiences of patients undertaking a 'virtual' cardiac rehabilitation program. *Stud Health Technol Inform* 2015;209:9-14. [Medline: [25980699](https://pubmed.ncbi.nlm.nih.gov/25980699/)]
47. Lear SA, Singer J, Banner-Lukaris D, Horvat D, Park JE, Bates J, et al. Improving access to cardiac rehabilitation using the internet: a randomized trial. *Stud Health Technol Inform* 2015;209:58-66. [Medline: [25980706](https://pubmed.ncbi.nlm.nih.gov/25980706/)]
48. Barton AB, Okorodudu DE, Bosworth HB, Crowley MJ. Clinical inertia in a randomized trial of telemedicine-based chronic disease management: lessons learned. *Telemed J E Health* 2018;24(10):742-748. [doi: [10.1089/tmj.2017.0184](https://doi.org/10.1089/tmj.2017.0184)] [Medline: [29341850](https://pubmed.ncbi.nlm.nih.gov/29341850/)]
49. Donohue JM, Belnap BH, Men A, He F, Roberts MS, Schulberg HC, et al. Twelve-month cost-effectiveness of telephone-delivered collaborative care for treating depression following CABG surgery: a randomized controlled trial. *Gen Hosp Psychiatry* 2014;36(5):453-459 [FREE Full text] [doi: [10.1016/j.genhosppsych.2014.05.012](https://doi.org/10.1016/j.genhosppsych.2014.05.012)] [Medline: [24973911](https://pubmed.ncbi.nlm.nih.gov/24973911/)]
50. Furber S, Butler L, Phongsavan P, Mark A, Bauman A. Randomised controlled trial of a pedometer-based telephone intervention to increase physical activity among cardiac patients not attending cardiac rehabilitation. *Patient Educ Couns* 2010;80(2):212-218. [doi: [10.1016/j.pec.2009.11.012](https://doi.org/10.1016/j.pec.2009.11.012)] [Medline: [20022201](https://pubmed.ncbi.nlm.nih.gov/20022201/)]
51. Ho YL, Yu JY, Lin YH, Chen YH, Huang CC, Hsu TP, 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;16(6):e145 [FREE Full text] [doi: [10.2196/jmir.3346](https://doi.org/10.2196/jmir.3346)] [Medline: [24915187](https://pubmed.ncbi.nlm.nih.gov/24915187/)]
52. Hwang R, Morris NR, Mandrusiak A, Bruning J, Peters R, Korczyk D, et al. Cost-utility analysis of home-based telerehabilitation compared with centre-based rehabilitation in patients with heart failure. *Heart Lung Circ* 2019;28(12):1795-1803. [doi: [10.1016/j.hlc.2018.11.010](https://doi.org/10.1016/j.hlc.2018.11.010)] [Medline: [30528811](https://pubmed.ncbi.nlm.nih.gov/30528811/)]
53. Hwang R, Bruning J, Morris NR, Mandrusiak A, Russell T. Home-based telerehabilitation is not inferior to a centre-based program in patients with chronic heart failure: a randomised trial. *J Physiother* 2017;63(2):101-107 [FREE Full text] [doi: [10.1016/j.jphys.2017.02.017](https://doi.org/10.1016/j.jphys.2017.02.017)] [Medline: [28336297](https://pubmed.ncbi.nlm.nih.gov/28336297/)]
54. Kidholm K, Rasmussen MK, Andreasen JJ, Hansen J, Nielsen G, Spindler H, et al. Cost-utility analysis of a cardiac telerehabilitation program: the teledialog project. *Telemed J E Health* 2016;22(7):553-563 [FREE Full text] [doi: [10.1089/tmj.2015.0194](https://doi.org/10.1089/tmj.2015.0194)] [Medline: [26713491](https://pubmed.ncbi.nlm.nih.gov/26713491/)]
55. Laustsen S, Oestergaard LG, van Tulder M, Hjortdal VE, Petersen AK. Telemonitored exercise-based cardiac rehabilitation improves physical capacity and health-related quality of life. *J Telemed Telecare* 2020;26(1-2):36-44. [doi: [10.1177/1357633X18792808](https://doi.org/10.1177/1357633X18792808)] [Medline: [30134780](https://pubmed.ncbi.nlm.nih.gov/30134780/)]
56. Mizukawa M, Moriyama M, Yamamoto H, Rahman MM, Naka M, Kitagawa T, et al. Nurse-led collaborative management using telemonitoring improves quality of life and prevention of rehospitalization in patients with heart failure. *Int Heart J* 2019;60(6):1293-1302 [FREE Full text] [doi: [10.1536/ihj.19-313](https://doi.org/10.1536/ihj.19-313)] [Medline: [31735786](https://pubmed.ncbi.nlm.nih.gov/31735786/)]
57. Nolan RP, Upshur RE, Lynn H, Crichton T, Rukholm E, Stewart DE, et al. Therapeutic benefit of preventive telehealth counseling in the Community Outreach Heart Health and Risk Reduction Trial. *Am J Cardiol* 2011;107(5):690-696. [doi: [10.1016/j.amjcard.2010.10.050](https://doi.org/10.1016/j.amjcard.2010.10.050)] [Medline: [21215382](https://pubmed.ncbi.nlm.nih.gov/21215382/)]
58. O'Neil A, Hawkes AL, Atherton JJ, Patrao TA, Sanderson K, Wolfe R, et al. Telephone-delivered health coaching improves anxiety outcomes after myocardial infarction: the 'ProActive Heart' trial. *Eur J Prev Cardiol* 2014;21(1):30-38. [doi: [10.1177/2047487312460515](https://doi.org/10.1177/2047487312460515)] [Medline: [22956633](https://pubmed.ncbi.nlm.nih.gov/22956633/)]
59. Hawkes AL, Patrao TA, Atherton J, Ware RS, Taylor CB, O'Neil A, et al. Effect of a telephone-delivered coronary heart disease secondary prevention program (proactive heart) on quality of life and health behaviours: primary outcomes of a randomised controlled trial. *Int J Behav Med* 2013;20(3):413-424. [doi: [10.1007/s12529-012-9250-5](https://doi.org/10.1007/s12529-012-9250-5)] [Medline: [23012159](https://pubmed.ncbi.nlm.nih.gov/23012159/)]
60. Ventura Marra M, Lilly CL, Nelson KR, Woofter DR, 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;11(2):229 [FREE Full text] [doi: [10.3390/nu11020229](https://doi.org/10.3390/nu11020229)] [Medline: [30678197](https://pubmed.ncbi.nlm.nih.gov/30678197/)]
61. Baque E, Barber L, Sakzewski L, Boyd RN. Randomized controlled trial of web-based multimodal therapy for children with acquired brain injury to improve gross motor capacity and performance. *Clin Rehabil* 2017;31(6):722-732. [doi: [10.1177/0269215516651980](https://doi.org/10.1177/0269215516651980)] [Medline: [27271374](https://pubmed.ncbi.nlm.nih.gov/27271374/)]
62. Cartwright M, Hirani SP, Rixon L, Beynon M, Doll H, Bower P, Whole Systems Demonstrator Evaluation Team. Effect of telehealth on quality of life and psychological outcomes over 12 months (Whole Systems Demonstrator telehealth questionnaire study): nested study of patient reported outcomes in a pragmatic, cluster randomised controlled trial. *BMJ* 2013;346:f653 [FREE Full text] [doi: [10.1136/bmj.f653](https://doi.org/10.1136/bmj.f653)] [Medline: [23444424](https://pubmed.ncbi.nlm.nih.gov/23444424/)]

63. Ellis TD, Cavanaugh JT, DeAngelis T, Hendron K, Thomas CA, Saint-Hilaire M, et al. Comparative effectiveness of mHealth-supported exercise compared with exercise alone for people with Parkinson disease: randomized controlled pilot study. *Phys Ther* 2019;99(2):203-216. [doi: [10.1093/ptj/pzy131](https://doi.org/10.1093/ptj/pzy131)] [Medline: [30715489](https://pubmed.ncbi.nlm.nih.gov/30715489/)]
64. Garbutt JM, Banister C, Highstein G, Sterkel R, Epstein J, Bruns J, et al. Telephone coaching for parents of children with asthma: impact and lessons learned. *Arch Pediatr Adolesc Med* 2010;164(7):625-630. [doi: [10.1001/archpediatrics.2010.91](https://doi.org/10.1001/archpediatrics.2010.91)] [Medline: [20603462](https://pubmed.ncbi.nlm.nih.gov/20603462/)]
65. Swerczek LM, Banister C, Bloomberg GR, Bruns JM, Epstein J, Highstein GR, et al. A telephone coaching intervention to improve asthma self-management behaviors. *Pediatr Nurs* 2013;39(3):125-145. [Medline: [23926751](https://pubmed.ncbi.nlm.nih.gov/23926751/)]
66. Gehrman P, Shah MT, Miles A, Kuna S, Godleski L. Feasibility of group cognitive-behavioral treatment of insomnia delivered by clinical video telehealth. *Telemed J E Health* 2016;22(12):1041-1046. [doi: [10.1089/tmj.2016.0032](https://doi.org/10.1089/tmj.2016.0032)] [Medline: [27286369](https://pubmed.ncbi.nlm.nih.gov/27286369/)]
67. Glaser E, Richard C, Lussier M. The impact of a patient web communication intervention on reaching treatment suggested guidelines for chronic diseases: a randomized controlled trial. *Patient Educ Couns* 2017;100(11):2062-2070. [doi: [10.1016/j.pec.2017.05.022](https://doi.org/10.1016/j.pec.2017.05.022)] [Medline: [28535926](https://pubmed.ncbi.nlm.nih.gov/28535926/)]
68. Patja K, Absetz P, Auvinen A, Tokola K, Kytö J, Oksman E, et al. Health coaching by telephony to support self-care in chronic diseases: clinical outcomes from The TERVA randomized controlled trial. *BMC Health Serv Res* 2012;12:147 [FREE Full text] [doi: [10.1186/1472-6963-12-147](https://doi.org/10.1186/1472-6963-12-147)] [Medline: [22682298](https://pubmed.ncbi.nlm.nih.gov/22682298/)]
69. Paul SS, Gardner T, Hubbard A, Johnson J, Canning CG, Leaver A, et al. Feasibility of the PHYZ X 2U program: a mobile and cloud-based outreach service to improve chronic disease outcomes in underserved rural communities. *Aust J Prim Health* 2019;25(6):539-546. [doi: [10.1071/PY19090](https://doi.org/10.1071/PY19090)] [Medline: [31708012](https://pubmed.ncbi.nlm.nih.gov/31708012/)]
70. Støme LN, Pripp AH, Kværner JS, Kvaerner KJ. Acceptability, usability and utility of a personalised application in promoting behavioural change in patients with osteoarthritis: a feasibility study in Norway. *BMJ Open* 2019;9(1):e021608 [FREE Full text] [doi: [10.1136/bmjopen-2018-021608](https://doi.org/10.1136/bmjopen-2018-021608)] [Medline: [30696666](https://pubmed.ncbi.nlm.nih.gov/30696666/)]
71. Warmington K, Flewelling C, Kennedy CA, Shupak R, Papachristos A, Jones C, et al. Telemedicine delivery of patient education in remote Ontario communities: feasibility of an Advanced Clinician Practitioner in Arthritis Care (ACPAC)-led inflammatory arthritis education program. *Open Access Rheumatol* 2017;9:11-19 [FREE Full text] [doi: [10.2147/OARRR.S122015](https://doi.org/10.2147/OARRR.S122015)] [Medline: [28280400](https://pubmed.ncbi.nlm.nih.gov/28280400/)]
72. Kennedy CA, Warmington K, Flewelling C, Shupak R, Papachristos A, Jones C, et al. A prospective comparison of telemedicine versus in-person delivery of an interprofessional education program for adults with inflammatory arthritis. *J Telemed Telecare* 2017;23(2):197-206. [doi: [10.1177/1357633X16635342](https://doi.org/10.1177/1357633X16635342)] [Medline: [26957111](https://pubmed.ncbi.nlm.nih.gov/26957111/)]
73. You A, Kawamoto J, Smith JP. A pharmacist-managed telemedicine clinic for hepatitis C care: a descriptive analysis. *J Telemed Telecare* 2014;20(2):99-101. [doi: [10.1177/1357633X13519043](https://doi.org/10.1177/1357633X13519043)] [Medline: [24414398](https://pubmed.ncbi.nlm.nih.gov/24414398/)]
74. Alcazar B, de Lucas P, Soriano JB, Fernández-Nistal A, Fuster A, González-Moro JM, et al. The evaluation of a remote support program on quality of life and evolution of disease in COPD patients with frequent exacerbations. *BMC Pulm Med* 2016;16(1):140 [FREE Full text] [doi: [10.1186/s12890-016-0304-3](https://doi.org/10.1186/s12890-016-0304-3)] [Medline: [27821164](https://pubmed.ncbi.nlm.nih.gov/27821164/)]
75. Antoniadis NC, Rochford PD, Pretto JJ, Pierce RJ, Gogler J, Steinkrug J, et al. Pilot study of remote telemonitoring in COPD. *Telemed J E Health* 2012;18(8):634-640. [doi: [10.1089/tmj.2011.0231](https://doi.org/10.1089/tmj.2011.0231)] [Medline: [22957501](https://pubmed.ncbi.nlm.nih.gov/22957501/)]
76. Barberan-Garcia A, Vogiatzis I, Solberg HS, Vilaró J, Rodríguez DA, Garåsen HM, NEXES Consortium. Effects and barriers to deployment of telehealth wellness programs for chronic patients across 3 European countries. *Respir Med* 2014;108(4):628-637 [FREE Full text] [doi: [10.1016/j.rmed.2013.12.006](https://doi.org/10.1016/j.rmed.2013.12.006)] [Medline: [24451438](https://pubmed.ncbi.nlm.nih.gov/24451438/)]
77. Barbita J, Neves-Silva S. Partnering with patients: the Toronto central LHIN telehomecare experience. *Healthc Q* 2017;20(3):36-40. [Medline: [29132448](https://pubmed.ncbi.nlm.nih.gov/29132448/)]
78. Blumenthal JA, Emery CF, Smith PJ, Keefe FJ, Welty-Wolf K, Mabe S, et al. The effects of a telehealth coping skills intervention on outcomes in chronic obstructive pulmonary disease: primary results from the INSPIRE-II study. *Psychosom Med* 2014;76(8):581-592 [FREE Full text] [doi: [10.1097/PSY.000000000000101](https://doi.org/10.1097/PSY.000000000000101)] [Medline: [25251888](https://pubmed.ncbi.nlm.nih.gov/25251888/)]
79. Broendum E, Ulrik CS, Gregersen T, Hansen EF, Green A, Ringbaek T. Barriers for recruitment of patients with chronic obstructive pulmonary disease to a controlled telemedicine trial. *Health Informatics J* 2018;24(2):216-224 [FREE Full text] [doi: [10.1177/1460458216667166](https://doi.org/10.1177/1460458216667166)] [Medline: [27638452](https://pubmed.ncbi.nlm.nih.gov/27638452/)]
80. Ringbæk T, Green A, Laursen LC, Frausing E, Brøndum E, Ulrik CS. Effect of tele health care on exacerbations and hospital admissions in patients with chronic obstructive pulmonary disease: a randomized clinical trial. *Int J Chron Obstruct Pulmon Dis* 2015;10:1801-1808 [FREE Full text] [doi: [10.2147/COPD.S85596](https://doi.org/10.2147/COPD.S85596)] [Medline: [26366072](https://pubmed.ncbi.nlm.nih.gov/26366072/)]
81. Cameron-Tucker HL, Wood-Baker R, Joseph L, Walters JA, Schüz N, Walters EH. A randomized controlled trial of telephone-mentoring with home-based walking preceding rehabilitation in COPD. *Int J Chron Obstruct Pulmon Dis* 2016;11:1991-2000 [FREE Full text] [doi: [10.2147/COPD.S109820](https://doi.org/10.2147/COPD.S109820)] [Medline: [27601892](https://pubmed.ncbi.nlm.nih.gov/27601892/)]
82. Esteban C, Moraza J, Iriberrri M, Aguirre U, Goiria B, Quintana JM, et al. Outcomes of a telemonitoring-based program (telePOC) in frequently hospitalized COPD patients. *Int J Chron Obstruct Pulmon Dis* 2016;11:2919-2930 [FREE Full text] [doi: [10.2147/COPD.S115350](https://doi.org/10.2147/COPD.S115350)] [Medline: [27920519](https://pubmed.ncbi.nlm.nih.gov/27920519/)]

83. Ure J, Pinnock H, Hanley J, Kidd G, McCall Smith E, Tarling A, et al. Piloting tele-monitoring in COPD: a mixed methods exploration of issues in design and implementation. *Prim Care Respir J* 2012;21(1):57-64 [[FREE Full text](#)] [doi: [10.4104/pcrj.2011.00065](https://doi.org/10.4104/pcrj.2011.00065)] [Medline: [21785816](https://pubmed.ncbi.nlm.nih.gov/21785816/)]
84. Fairbrother P, Pinnock H, Hanley J, McCloughan L, Sheikh A, Pagliari C, TELESCOT programme team. Continuity, but at what cost? The impact of telemonitoring COPD on continuities of care: a qualitative study. *Prim Care Respir J* 2012;21(3):322-328 [[FREE Full text](#)] [doi: [10.4104/pcrj.2012.00068](https://doi.org/10.4104/pcrj.2012.00068)] [Medline: [22875143](https://pubmed.ncbi.nlm.nih.gov/22875143/)]
85. Pinnock H, Hanley J, McCloughan L, Todd A, Krishan A, Lewis S, et al. Effectiveness of telemonitoring integrated into existing clinical services on hospital admission for exacerbation of chronic obstructive pulmonary disease: researcher blind, multicentre, randomised controlled trial. *BMJ* 2013;347:f6070 [[FREE Full text](#)] [doi: [10.1136/bmj.f6070](https://doi.org/10.1136/bmj.f6070)] [Medline: [24136634](https://pubmed.ncbi.nlm.nih.gov/24136634/)]
86. Kessler R, Casan-Clara P, Koehler D, Tognella S, Viejo JL, Dal Negro RW, et al. COMET: a multicomponent home-based disease-management programme versus routine care in severe COPD. *Eur Respir J* 2018;51(1):1701612 [[FREE Full text](#)] [doi: [10.1183/13993003.01612-2017](https://doi.org/10.1183/13993003.01612-2017)] [Medline: [29326333](https://pubmed.ncbi.nlm.nih.gov/29326333/)]
87. Kim J, Kim S, Kim HC, Kim KH, Yang SC, Lee CT, et al. Effects of consumer-centered u-health service for the knowledge, skill, and attitude of the patients with chronic obstructive pulmonary disease. *Comput Inform Nurs* 2012;30(12):661-671. [doi: [10.1097/NXN.0b013e318261c1c](https://doi.org/10.1097/NXN.0b013e318261c1c)] [Medline: [23266537](https://pubmed.ncbi.nlm.nih.gov/23266537/)]
88. Nissen L, Lindhardt T. A qualitative study of COPD-patients' experience of a telemedicine intervention. *Int J Med Inform* 2017;107:11-17. [doi: [10.1016/j.ijmedinf.2017.08.004](https://doi.org/10.1016/j.ijmedinf.2017.08.004)] [Medline: [29029687](https://pubmed.ncbi.nlm.nih.gov/29029687/)]
89. Tabak M, Brusse-Keizer M, van der Valk P, Hermens H, Vollenbroek-Hutten M. A telehealth program for self-management of COPD exacerbations and promotion of an active lifestyle: a pilot randomized controlled trial. *Int J Chron Obstruct Pulmon Dis* 2014;9:935-944 [[FREE Full text](#)] [doi: [10.2147/COPD.S60179](https://doi.org/10.2147/COPD.S60179)] [Medline: [25246781](https://pubmed.ncbi.nlm.nih.gov/25246781/)]
90. Zanaboni P, Lien LA, Hjalmsen A, Wootton R. Long-term telerehabilitation of COPD patients in their homes: interim results from a pilot study in Northern Norway. *J Telemed Telecare* 2013;19(7):425-429. [doi: [10.1177/1357633X13506514](https://doi.org/10.1177/1357633X13506514)] [Medline: [24218358](https://pubmed.ncbi.nlm.nih.gov/24218358/)]
91. Bennell KL, Nelligan R, Dobson F, Rini C, Keefe F, Kasza J, et al. Effectiveness of an internet-delivered exercise and pain-coping skills training intervention for persons with chronic knee pain: a randomized trial. *Ann Intern Med* 2017;166(7):453-462. [doi: [10.7326/M16-1714](https://doi.org/10.7326/M16-1714)] [Medline: [28241215](https://pubmed.ncbi.nlm.nih.gov/28241215/)]
92. Williams A, Wiggers J, O'Brien KM, Wolfenden L, Yoong SL, Hodder RK, et al. Effectiveness of a healthy lifestyle intervention for chronic low back pain: a randomised controlled trial. *Pain* 2018;159(6):1137-1146. [doi: [10.1097/j.pain.0000000000001198](https://doi.org/10.1097/j.pain.0000000000001198)] [Medline: [29528963](https://pubmed.ncbi.nlm.nih.gov/29528963/)]
93. Baron J, Hirani S, Newman S. Challenges in patient recruitment, implementation, and fidelity in a mobile telehealth study. *Telemed J E Health* 2016;22(5):400-409. [doi: [10.1089/tmj.2015.0095](https://doi.org/10.1089/tmj.2015.0095)] [Medline: [26540238](https://pubmed.ncbi.nlm.nih.gov/26540238/)]
94. Baron JS, Hirani SP, Newman SP. Investigating the behavioural effects of a mobile-phone based home telehealth intervention in people with insulin-requiring diabetes: results of a randomized controlled trial with patient interviews. *J Telemed Telecare* 2017;23(5):503-512. [doi: [10.1177/1357633X16655911](https://doi.org/10.1177/1357633X16655911)] [Medline: [27377790](https://pubmed.ncbi.nlm.nih.gov/27377790/)]
95. Baron JS, Hirani S, Newman SP. A randomised, controlled trial of the effects of a mobile telehealth intervention on clinical and patient-reported outcomes in people with poorly controlled diabetes. *J Telemed Telecare* 2017;23(2):207-216. [doi: [10.1177/1357633X16631628](https://doi.org/10.1177/1357633X16631628)] [Medline: [26880694](https://pubmed.ncbi.nlm.nih.gov/26880694/)]
96. Blackberry ID, Furler JS, Best JD, Chondros P, Vale M, Walker C, et al. Effectiveness of general practice based, practice nurse led telephone coaching on glycaemic control of type 2 diabetes: the Patient Engagement and Coaching for Health (PEACH) pragmatic cluster randomised controlled trial. *BMJ* 2013;347:f5272 [[FREE Full text](#)] [doi: [10.1136/bmj.f5272](https://doi.org/10.1136/bmj.f5272)] [Medline: [24048296](https://pubmed.ncbi.nlm.nih.gov/24048296/)]
97. Buysse H, Coremans P, Pouwer F, Ruige J. Sustainable improvement of HbA1c and satisfaction with diabetes care after adding telemedicine in patients on adaptable insulin regimens: results of the TeleDiabetes randomized controlled trial. *Health Informatics J* 2020;26(1):628-641 [[FREE Full text](#)] [doi: [10.1177/1460458219844369](https://doi.org/10.1177/1460458219844369)] [Medline: [31046527](https://pubmed.ncbi.nlm.nih.gov/31046527/)]
98. Carter EL, Nunlee-Bland G, Callender C. A patient-centric, provider-assisted diabetes telehealth self-management intervention for urban minorities. *Perspect Health Inf Manag* 2011;8(Winter):1b [[FREE Full text](#)] [Medline: [21307985](https://pubmed.ncbi.nlm.nih.gov/21307985/)]
99. Charpentier G, Benhamou PY, Dardari D, Clergeot A, Franc S, Schaepelynck-Belicar P, TeleDiab Study Group. The Diabeo software enabling individualized insulin dose adjustments combined with telemedicine support improves HbA1c in poorly controlled type 1 diabetic patients: a 6-month, randomized, open-label, parallel-group, multicenter trial (TeleDiab 1 Study). *Diabetes Care* 2011;34(3):533-539 [[FREE Full text](#)] [doi: [10.2337/dc10-1259](https://doi.org/10.2337/dc10-1259)] [Medline: [21266648](https://pubmed.ncbi.nlm.nih.gov/21266648/)]
100. Chiu CJ, Yu YC, Du YF, Yang YC, Chen JY, Wong LP, et al. Comparing a social and communication app, telephone intervention, and usual care for diabetes self-management: 3-arm Quasiexperimental evaluation study. *JMIR Mhealth Uhealth* 2020;8(6):e14024 [[FREE Full text](#)] [doi: [10.2196/14024](https://doi.org/10.2196/14024)] [Medline: [32484448](https://pubmed.ncbi.nlm.nih.gov/32484448/)]
101. Ciemins EL, Coon PJ, Coombs NC, Holloway BL, Mullette EJ, Dudley WN. Intent-to-treat analysis of a simultaneous multisite telehealth diabetes prevention program. *BMJ Open Diabetes Res Care* 2018;6(1):e000515 [[FREE Full text](#)] [doi: [10.1136/bmjdr-2018-000515](https://doi.org/10.1136/bmjdr-2018-000515)] [Medline: [29713481](https://pubmed.ncbi.nlm.nih.gov/29713481/)]

102. Davis RM, Hitch AD, Salaam MM, Herman WH, Zimmer-Galler IE, Mayer-Davis EJ. TeleHealth improves diabetes self-management in an underserved community: diabetes TeleCare. *Diabetes Care* 2010;33(8):1712-1717 [[FREE Full text](#)] [doi: [10.2337/dc09-1919](https://doi.org/10.2337/dc09-1919)] [Medline: [20484125](https://pubmed.ncbi.nlm.nih.gov/20484125/)]
103. du Pon E, Kleefstra N, Cleveringa F, van Dooren A, Heerdink ER, van Dulmen S. Effects of the proactive interdisciplinary self-management (PRISMA) program on online care platform usage in patients with type 2 diabetes in primary care: a randomized controlled trial. *J Diabetes Res* 2020;2020:5013142 [[FREE Full text](#)] [doi: [10.1155/2020/5013142](https://doi.org/10.1155/2020/5013142)] [Medline: [32016122](https://pubmed.ncbi.nlm.nih.gov/32016122/)]
104. Eakin EG, Winkler EA, Dunstan DW, Healy GN, Owen N, Marshall AM, et al. Living well with diabetes: 24-month outcomes from a randomized trial of telephone-delivered weight loss and physical activity intervention to improve glycemic control. *Diabetes Care* 2014;37(8):2177-2185. [doi: [10.2337/dc13-2427](https://doi.org/10.2337/dc13-2427)] [Medline: [24658390](https://pubmed.ncbi.nlm.nih.gov/24658390/)]
105. Goode AD, Winkler EA, Reeves MM, Eakin EG. Relationship between intervention dose and outcomes in living well with diabetes--a randomized trial of a telephone-delivered lifestyle-based weight loss intervention. *Am J Health Promot* 2015;30(2):120-129. [doi: [10.4278/ajhp.140206-QUAN-62](https://doi.org/10.4278/ajhp.140206-QUAN-62)] [Medline: [25372235](https://pubmed.ncbi.nlm.nih.gov/25372235/)]
106. Esmatjes E, Jansà M, Roca D, Pérez-Ferre N, del Valle L, Martínez-Hervás S, Telemed-Diabetes Group. The efficiency of telemedicine to optimize metabolic control in patients with type 1 diabetes mellitus: Telemed study. *Diabetes Technol Ther* 2014;16(7):435-441. [doi: [10.1089/dia.2013.0313](https://doi.org/10.1089/dia.2013.0313)] [Medline: [24528195](https://pubmed.ncbi.nlm.nih.gov/24528195/)]
107. Fountoulakis S, Papanastasiou L, Gryparis A, Markou A, Piaditis G. Impact and duration effect of telemonitoring on HbA1c, BMI and cost in insulin-treated diabetes mellitus patients with inadequate glycemic control: a randomized controlled study. *Hormones (Athens)* 2015;14(4):632-643 [[FREE Full text](#)] [doi: [10.14310/horm.2002.1603](https://doi.org/10.14310/horm.2002.1603)] [Medline: [26188234](https://pubmed.ncbi.nlm.nih.gov/26188234/)]
108. Greenwood DA, Blozis SA, Young HM, Nesbitt TS, Quinn CC. Overcoming clinical inertia: a randomized clinical trial of a telehealth remote monitoring intervention using paired glucose testing in adults with type 2 diabetes. *J Med Internet Res* 2015;17(7):e178 [[FREE Full text](#)] [doi: [10.2196/jmir.4112](https://doi.org/10.2196/jmir.4112)] [Medline: [26199142](https://pubmed.ncbi.nlm.nih.gov/26199142/)]
109. Hanley J, Fairbrother P, McCloughan L, Pagliari C, Paterson M, Pinnock H, et al. Qualitative study of telemonitoring of blood glucose and blood pressure in type 2 diabetes. *BMJ Open* 2015;5(12):e008896 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2015-008896](https://doi.org/10.1136/bmjopen-2015-008896)] [Medline: [26700275](https://pubmed.ncbi.nlm.nih.gov/26700275/)]
110. Wild SH, Hanley J, Lewis SC, McKnight JA, McCloughan LB, Padfield PL, et al. Supported telemonitoring and glycemic control in people with type 2 diabetes: the telescot diabetes pragmatic multicenter randomized controlled trial. *PLoS Med* 2016;13(7):e1002098 [[FREE Full text](#)] [doi: [10.1371/journal.pmed.1002098](https://doi.org/10.1371/journal.pmed.1002098)] [Medline: [27458809](https://pubmed.ncbi.nlm.nih.gov/27458809/)]
111. Trief PM, Izquierdo R, Eimicke JP, Teresi JA, Goland R, Palmas W, et al. Adherence to diabetes self care for White, African-American and Hispanic American telemedicine participants: 5 year results from the IDEATel project. *Ethn Health* 2013;18(1):83-96. [doi: [10.1080/13557858.2012.700915](https://doi.org/10.1080/13557858.2012.700915)] [Medline: [22762449](https://pubmed.ncbi.nlm.nih.gov/22762449/)]
112. Izquierdo R, Laguna CT, Meyer S, Ploutz-Snyder RJ, Palmas W, Eimicke JP, et al. Telemedicine intervention effects on waist circumference and body mass index in the IDEATel project. *Diabetes Technol Ther* 2010;12(3):213-220. [doi: [10.1089/dia.2009.0102](https://doi.org/10.1089/dia.2009.0102)] [Medline: [20151772](https://pubmed.ncbi.nlm.nih.gov/20151772/)]
113. Weinstock RS, Teresi JA, Goland R, Izquierdo R, Palmas W, Eimicke JP, IDEATel Consortium. Glycemic control and health disparities in older ethnically diverse underserved adults with diabetes: five-year results from the Informatics for Diabetes Education and Telemedicine (IDEATel) study. *Diabetes Care* 2011;34(2):274-279 [[FREE Full text](#)] [doi: [10.2337/dc10-1346](https://doi.org/10.2337/dc10-1346)] [Medline: [21270184](https://pubmed.ncbi.nlm.nih.gov/21270184/)]
114. Weinstock RS, Brooks G, Palmas W, Morin PC, Teresi JA, Eimicke JP, et al. Lessened decline in physical activity and impairment of older adults with diabetes with telemedicine and pedometer use: results from the IDEATel study. *Age Ageing* 2011;40(1):98-105. [doi: [10.1093/ageing/afq147](https://doi.org/10.1093/ageing/afq147)] [Medline: [21081539](https://pubmed.ncbi.nlm.nih.gov/21081539/)]
115. West SP, Laguna C, Trief PM, Izquierdo R, Weinstock RS. Goal setting using telemedicine in rural underserved older adults with diabetes: experiences from the informatics for diabetes education and telemedicine project. *Telemed J E Health* 2010;16(4):405-416. [doi: [10.1089/tmj.2009.0136](https://doi.org/10.1089/tmj.2009.0136)] [Medline: [20507198](https://pubmed.ncbi.nlm.nih.gov/20507198/)]
116. Kearns JW, Bowerman D, Kemmis K, Izquierdo RE, Wade M, Weinstock RS. Group diabetes education administered through telemedicine: tools used and lessons learned. *Telemed J E Health* 2012;18(5):347-353. [doi: [10.1089/tmj.2011.0165](https://doi.org/10.1089/tmj.2011.0165)] [Medline: [22468984](https://pubmed.ncbi.nlm.nih.gov/22468984/)]
117. Klingeman H, Funnell M, Jhand A, Lathkar-Pradhan S, Hodish I. Type 2 diabetes specialty clinic model for the accountable care organization era. *J Diabetes Complications* 2017;31(10):1521-1526. [doi: [10.1016/j.jdiacomp.2017.05.011](https://doi.org/10.1016/j.jdiacomp.2017.05.011)] [Medline: [28793967](https://pubmed.ncbi.nlm.nih.gov/28793967/)]
118. Lyles CR, Harris LT, Le T, Flowers J, Tufano J, Britt D, et al. Qualitative evaluation of a mobile phone and web-based collaborative care intervention for patients with type 2 diabetes. *Diabetes Technol Ther* 2011;13(5):563-569. [doi: [10.1089/dia.2010.0200](https://doi.org/10.1089/dia.2010.0200)] [Medline: [21406018](https://pubmed.ncbi.nlm.nih.gov/21406018/)]
119. McLendon SF, Wood FG, Stanley N. Enhancing diabetes care through care coordination, telemedicine, and education: evaluation of a rural pilot program. *Public Health Nurs* 2019;36(3):310-320. [doi: [10.1111/phn.12601](https://doi.org/10.1111/phn.12601)] [Medline: [30868661](https://pubmed.ncbi.nlm.nih.gov/30868661/)]
120. Musacchio N, Lovagnini Scher A, Giancaterini A, Pessina L, Salis G, Schivalocchi F, et al. Impact of a chronic care model based on patient empowerment on the management of type 2 diabetes: effects of the SINERGIA programme. *Diabet Med* 2011;28(6):724-730. [doi: [10.1111/j.1464-5491.2011.03253.x](https://doi.org/10.1111/j.1464-5491.2011.03253.x)] [Medline: [21294769](https://pubmed.ncbi.nlm.nih.gov/21294769/)]

121. Pludwinski S, Ahmad F, Wayne N, Ritvo P. Participant experiences in a smartphone-based health coaching intervention for type 2 diabetes: a qualitative inquiry. *J Telemed Telecare* 2016;22(3):172-178. [doi: [10.1177/1357633X15595178](https://doi.org/10.1177/1357633X15595178)] [Medline: [26199275](https://pubmed.ncbi.nlm.nih.gov/26199275/)]
122. Reinhardt JA, van der Ploeg HP, Grzegorzulka R, Timperley JG. Implementing lifestyle change through phone-based motivational interviewing in rural-based women with previous gestational diabetes mellitus. *Health Promot J Austr* 2012;23(1):5-9. [doi: [10.1071/he12005](https://doi.org/10.1071/he12005)] [Medline: [22730940](https://pubmed.ncbi.nlm.nih.gov/22730940/)]
123. Ryan JG, Schwartz R, Jennings T, Fedders M, Vittoria I. Feasibility of an internet-based intervention for improving diabetes outcomes among low-income patients with a high risk for poor diabetes outcomes followed in a community clinic. *Diabetes Educ* 2013;39(3):365-375. [doi: [10.1177/0145721713484594](https://doi.org/10.1177/0145721713484594)] [Medline: [23610181](https://pubmed.ncbi.nlm.nih.gov/23610181/)]
124. Sood A, Watts SA, Johnson JK, Hirth S, Aron DC. Telemedicine consultation for patients with diabetes mellitus: a cluster randomised controlled trial. *J Telemed Telecare* 2018;24(6):385-391. [doi: [10.1177/1357633X17704346](https://doi.org/10.1177/1357633X17704346)] [Medline: [28406066](https://pubmed.ncbi.nlm.nih.gov/28406066/)]
125. Vadheim LM, McPherson C, Kassner DR, Vanderwood KK, Hall TO, Butcher MK, et al. Adapted diabetes prevention program lifestyle intervention can be effectively delivered through telehealth. *Diabetes Educ* 2010;36(4):651-656. [doi: [10.1177/0145721710372811](https://doi.org/10.1177/0145721710372811)] [Medline: [20534873](https://pubmed.ncbi.nlm.nih.gov/20534873/)]
126. Van Dyck D, De Greef K, Deforche B, Ruige J, Bouckaert J, Tudor-Locke CE, et al. The relationship between changes in steps/day and health outcomes after a pedometer-based physical activity intervention with telephone support in type 2 diabetes patients. *Health Educ Res* 2013;28(3):539-545. [doi: [10.1093/her/cyt038](https://doi.org/10.1093/her/cyt038)] [Medline: [23492248](https://pubmed.ncbi.nlm.nih.gov/23492248/)]
127. Varney JE, Weiland TJ, Inder WJ, Jelinek GA. Effect of hospital-based telephone coaching on glycaemic control and adherence to management guidelines in type 2 diabetes, a randomised controlled trial. *Intern Med J* 2014;44(9):890-897. [doi: [10.1111/imj.12515](https://doi.org/10.1111/imj.12515)] [Medline: [24963611](https://pubmed.ncbi.nlm.nih.gov/24963611/)]
128. Wai Leng C, Jundong J, Li Wei C, Joo Pin F, Kwong Ming F, Chen R. Telehealth for improved glycaemic control in patients with poorly controlled diabetes after acute hospitalization - a preliminary study in Singapore. *J Telemed Telecare* 2014;20(6):317-323. [doi: [10.1177/1357633X14544441](https://doi.org/10.1177/1357633X14544441)] [Medline: [25052389](https://pubmed.ncbi.nlm.nih.gov/25052389/)]
129. Wakefield BJ, Holman JE, Ray A, Scherubel M, Adams MR, Hillis SL, et al. Effectiveness of home telehealth in comorbid diabetes and hypertension: a randomized, controlled trial. *Telemed J E Health* 2011;17(4):254-261. [doi: [10.1089/tmj.2010.0176](https://doi.org/10.1089/tmj.2010.0176)] [Medline: [21476945](https://pubmed.ncbi.nlm.nih.gov/21476945/)]
130. Whitehouse CR, Long JA, Maloney LM, Daniels K, Horowitz DA, Bowles KH. Feasibility of diabetes self-management telehealth education for older adults during transitions in care. *Res Gerontol Nurs* 2020;13(3):138-145 [FREE Full text] [doi: [10.3928/19404921-20191210-03](https://doi.org/10.3928/19404921-20191210-03)] [Medline: [31834415](https://pubmed.ncbi.nlm.nih.gov/31834415/)]
131. Ahrendt AD, Kattelman KK, Rector TS, Maddox DA. The effectiveness of telemedicine for weight management in the MOVE! program. *J Rural Health* 2014;30(1):113-119. [doi: [10.1111/jrh.12049](https://doi.org/10.1111/jrh.12049)] [Medline: [24112223](https://pubmed.ncbi.nlm.nih.gov/24112223/)]
132. Barrett S, Begg S, O'Halloran P, Kingsley M. Cost-effectiveness of telephone coaching for physically inactive ambulatory care hospital patients: economic evaluation alongside the Healthy4U randomised controlled trial. *BMJ Open* 2019;9(12):e032500 [FREE Full text] [doi: [10.1136/bmjopen-2019-032500](https://doi.org/10.1136/bmjopen-2019-032500)] [Medline: [31826893](https://pubmed.ncbi.nlm.nih.gov/31826893/)]
133. Cox CE, Hough CL, Carson SS, White DB, Kahn JM, Olsen MK, et al. Effects of a telephone- and web-based coping skills training program compared with an education program for survivors of critical illness and their family members. A randomized clinical trial. *Am J Respir Crit Care Med* 2018;197(1):66-78. [doi: [10.1164/rccm.201704-0720OC](https://doi.org/10.1164/rccm.201704-0720OC)] [Medline: [28872898](https://pubmed.ncbi.nlm.nih.gov/28872898/)]
134. Donaldson EL, Fallows S, Morris M. A text message based weight management intervention for overweight adults. *J Hum Nutr Diet* 2014;27 Suppl 2:90-97. [doi: [10.1111/jhn.12096](https://doi.org/10.1111/jhn.12096)] [Medline: [23738786](https://pubmed.ncbi.nlm.nih.gov/23738786/)]
135. Ferrara A, Hedderson MM, Brown SD, Ehrlich SF, Tsai A, Feng J, et al. A telehealth lifestyle intervention to reduce excess gestational weight gain in pregnant women with overweight or obesity (GLOW): a randomised, parallel-group, controlled trial. *Lancet Diabetes Endocrinol* 2020;8(6):490-500. [doi: [10.1016/S2213-8587\(20\)30107-8](https://doi.org/10.1016/S2213-8587(20)30107-8)] [Medline: [32445736](https://pubmed.ncbi.nlm.nih.gov/32445736/)]
136. 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;7(4):e12289 [FREE Full text] [doi: [10.2196/12289](https://doi.org/10.2196/12289)] [Medline: [30973338](https://pubmed.ncbi.nlm.nih.gov/30973338/)]
137. Han HR, Kim J, Kim KB, Jeong S, Levine D, Li C, et al. Implementation and success of nurse telephone counseling in linguistically isolated Korean American patients with high blood pressure. *Patient Educ Couns* 2010;80(1):130-134 [FREE Full text] [doi: [10.1016/j.pec.2009.10.012](https://doi.org/10.1016/j.pec.2009.10.012)] [Medline: [19945816](https://pubmed.ncbi.nlm.nih.gov/19945816/)]
138. Padwal RS, Klarenbach S, Sharma AM, Fradette M, Jelinski SE, Edwards A, et al. The evaluating self-management and educational support in severely obese patients awaiting multidisciplinary bariatric care (EVOLUTION) trial: principal results. *BMC Med* 2017;15(1):46 [FREE Full text] [doi: [10.1186/s12916-017-0808-6](https://doi.org/10.1186/s12916-017-0808-6)] [Medline: [28249576](https://pubmed.ncbi.nlm.nih.gov/28249576/)]
139. Rossi MC, Perozzi C, Consorti C, Almonti T, Fogliani P, Giostra N, et al. An interactive diary for diet management (DAI): a new telemedicine system able to promote body weight reduction, nutritional education, and consumption of fresh local produce. *Diabetes Technol Ther* 2010;12(8):641-647. [doi: [10.1089/dia.2010.0025](https://doi.org/10.1089/dia.2010.0025)] [Medline: [20615106](https://pubmed.ncbi.nlm.nih.gov/20615106/)]
140. Schweickert PA, Rutledge CM, Cattell-Gordon DC, Solenski NJ, Jensen ME, Branson S, et al. Telehealth stroke education for rural elderly Virginians. *Telemed J E Health* 2011;17(10):784-788. [doi: [10.1089/tmj.2011.0080](https://doi.org/10.1089/tmj.2011.0080)] [Medline: [22011051](https://pubmed.ncbi.nlm.nih.gov/22011051/)]

141. Weinstock RS, Trief PM, Cibula D, Morin PC, Delahanty LM. Weight loss success in metabolic syndrome by telephone interventions: results from the SHINE Study. *J Gen Intern Med* 2013;28(12):1620-1628 [FREE Full text] [doi: [10.1007/s11606-013-2529-7](https://doi.org/10.1007/s11606-013-2529-7)] [Medline: [23843020](https://pubmed.ncbi.nlm.nih.gov/23843020/)]
142. Data for high income, OECD members, upper middle income. The World Bank. 2020. URL: <https://data.worldbank.org/?locations=XD-OE-XT> [accessed 2020-10-30]
143. Aromataris E, Munn Z. JBI systematic reviews. In: Aromataris E, Munn Z, editors. *JBI Manual for Evidence synthesis*. Adelaide, Australia: Joanna Briggs Institute; 2017.
144. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 2018;18(1):143 [FREE Full text] [doi: [10.1186/s12874-018-0611-x](https://doi.org/10.1186/s12874-018-0611-x)] [Medline: [30453902](https://pubmed.ncbi.nlm.nih.gov/30453902/)]
145. Wagner EH, Austin BT, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: translating evidence into action. *Health Aff (Millwood)* 2001;20(6):64-78. [doi: [10.1377/hlthaff.20.6.64](https://doi.org/10.1377/hlthaff.20.6.64)] [Medline: [11816692](https://pubmed.ncbi.nlm.nih.gov/11816692/)]
146. Prochaska JO, DiClemente CC, Norcross JC. In search of how people change. Applications to addictive behaviors. *Am Psychol* 1992 Sep;47(9):1102-1114. [doi: [10.1037//0003-066x.47.9.1102](https://doi.org/10.1037//0003-066x.47.9.1102)] [Medline: [1329589](https://pubmed.ncbi.nlm.nih.gov/1329589/)]
147. Health at a glance 2019: OECD indicators. Organisation for Economic Co-operation and Development. 2019. URL: <http://www.oecd.org/health/health-systems/health-at-a-glance-19991312.htm> [accessed 2020-11-04]
148. Safavi KC, Cohen AB, Ting DY, Chaguturu S, Rowe JS. Health systems as venture capital investors in digital health: 2011-2019. *NPJ Digit Med* 2020;3:103 [FREE Full text] [doi: [10.1038/s41746-020-00311-5](https://doi.org/10.1038/s41746-020-00311-5)] [Medline: [32802968](https://pubmed.ncbi.nlm.nih.gov/32802968/)]
149. Lewis T, Synowiec C, Lagomarsino G, Schweitzer J. E-health in low- and middle-income countries: findings from the Center for Health Market Innovations. *Bull World Health Organ* 2012;90(5):332-340 [FREE Full text] [doi: [10.2471/BLT.11.099820](https://doi.org/10.2471/BLT.11.099820)] [Medline: [22589566](https://pubmed.ncbi.nlm.nih.gov/22589566/)]
150. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. *Lancet Psychiatry* 2017;4(6):486-500 [FREE Full text] [doi: [10.1016/S2215-0366\(17\)30096-2](https://doi.org/10.1016/S2215-0366(17)30096-2)] [Medline: [28433615](https://pubmed.ncbi.nlm.nih.gov/28433615/)]
151. Sondaal SF, Browne JL, Amoakoh-Coleman M, Borgstein A, Miltenburg AS, Verwijs M, et al. Assessing the effect of mHealth interventions in improving maternal and neonatal care in low- and middle-income countries: a systematic review. *PLoS One* 2016;11(5):e0154664 [FREE Full text] [doi: [10.1371/journal.pone.0154664](https://doi.org/10.1371/journal.pone.0154664)] [Medline: [27144393](https://pubmed.ncbi.nlm.nih.gov/27144393/)]
152. Peek N, Suján M, Scott P. Digital health and care in pandemic times: impact of COVID-19. *BMJ Health Care Inform* 2020;27(1):e100166 [FREE Full text] [doi: [10.1136/bmjhci-2020-100166](https://doi.org/10.1136/bmjhci-2020-100166)] [Medline: [32565418](https://pubmed.ncbi.nlm.nih.gov/32565418/)]
153. 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;392(10159):1789-1858 [FREE Full text] [doi: [10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7)] [Medline: [30496104](https://pubmed.ncbi.nlm.nih.gov/30496104/)]
154. Blüher M. Obesity: global epidemiology and pathogenesis. *Nat Rev Endocrinol* 2019;15(5):288-298. [doi: [10.1038/s41574-019-0176-8](https://doi.org/10.1038/s41574-019-0176-8)] [Medline: [30814686](https://pubmed.ncbi.nlm.nih.gov/30814686/)]
155. De Rosis S, Seghieri C. Basic ICT adoption and use by general practitioners: an analysis of primary care systems in 31 European countries. *BMC Med Inform Decis Mak* 2015;15:70 [FREE Full text] [doi: [10.1186/s12911-015-0185-z](https://doi.org/10.1186/s12911-015-0185-z)] [Medline: [26296994](https://pubmed.ncbi.nlm.nih.gov/26296994/)]
156. Moreno C, Wykes T, Galderisi S, Nordentoft M, Crossley N, Jones N, et al. How mental health care should change as a consequence of the COVID-19 pandemic. *Lancet Psychiatry* 2020;7(9):813-824 [FREE Full text] [doi: [10.1016/S2215-0366\(20\)30307-2](https://doi.org/10.1016/S2215-0366(20)30307-2)] [Medline: [32682460](https://pubmed.ncbi.nlm.nih.gov/32682460/)]
157. Dickinson C, Dow J, Gibson G, Hayes L, Robalino S, Robinson L. Psychosocial intervention for carers of people with dementia: what components are most effective and when? A systematic review of systematic reviews. *Int Psychogeriatr* 2017;29(1):31-43. [doi: [10.1017/S1041610216001447](https://doi.org/10.1017/S1041610216001447)] [Medline: [27666669](https://pubmed.ncbi.nlm.nih.gov/27666669/)]
158. Petkari E, Kaselionyte J, Altun S, Giacco D. Involvement of informal carers in discharge planning and transition between hospital and community mental health care: a systematic review. *J Psychiatr Ment Health Nurs* 2021;28(4):521-530. [doi: [10.1111/jpm.12701](https://doi.org/10.1111/jpm.12701)] [Medline: [33053271](https://pubmed.ncbi.nlm.nih.gov/33053271/)]
159. Heynsbergh N, Heckel L, Botti M, Livingston PM. Feasibility, useability and acceptability of technology-based interventions for informal cancer carers: a systematic review. *BMC Cancer* 2018;18(1):244 [FREE Full text] [doi: [10.1186/s12885-018-4160-9](https://doi.org/10.1186/s12885-018-4160-9)] [Medline: [29499663](https://pubmed.ncbi.nlm.nih.gov/29499663/)]
160. Snoswell CL, Caffery LJ, Haydon HM, Thomas EE, Smith AC. Telehealth uptake in general practice as a result of the coronavirus (COVID-19) pandemic. *Aust Health Rev* 2020;44(5):737-740. [doi: [10.1071/AH20183](https://doi.org/10.1071/AH20183)] [Medline: [32853536](https://pubmed.ncbi.nlm.nih.gov/32853536/)]
161. Gentry MT, Lapid MI, Clark MM, Rummans TA. Evidence for telehealth group-based treatment: a systematic review. *J Telemed Telecare* 2019;25(6):327-342. [doi: [10.1177/1357633X18775855](https://doi.org/10.1177/1357633X18775855)] [Medline: [29788807](https://pubmed.ncbi.nlm.nih.gov/29788807/)]
162. Hoey LM, Ieropoli SC, White VM, Jefford M. Systematic review of peer-support programs for people with cancer. *Patient Educ Couns* 2008;70(3):315-337. [doi: [10.1016/j.pec.2007.11.016](https://doi.org/10.1016/j.pec.2007.11.016)] [Medline: [18191527](https://pubmed.ncbi.nlm.nih.gov/18191527/)]
163. 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;20(5):e10235 [FREE Full text] [doi: [10.2196/10235](https://doi.org/10.2196/10235)] [Medline: [29716883](https://pubmed.ncbi.nlm.nih.gov/29716883/)]

164. Speyer R, Denman D, Wilkes-Gillan S, Chen YW, Bogaardt H, Kim JH, et al. Effects of telehealth by allied health professionals and nurses in rural and remote areas: a systematic review and meta-analysis. *J Rehabil Med* 2018;50(3):225-235 [[FREE Full text](#)] [doi: [10.2340/16501977-2297](https://doi.org/10.2340/16501977-2297)] [Medline: [29257195](https://pubmed.ncbi.nlm.nih.gov/29257195/)]
165. Luo C, Sanger N, Singhal N, Patrick K, Shams I, Shahid H, et al. A comparison of electronically-delivered and face to face cognitive behavioural therapies in depressive disorders: a systematic review and meta-analysis. *EClinicalMedicine* 2020;24:100442 [[FREE Full text](#)] [doi: [10.1016/j.eclinm.2020.100442](https://doi.org/10.1016/j.eclinm.2020.100442)] [Medline: [32775969](https://pubmed.ncbi.nlm.nih.gov/32775969/)]
166. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. *Curr Nutr Rep* 2016;5(3):160-167. [doi: [10.1007/s13668-016-0165-7](https://doi.org/10.1007/s13668-016-0165-7)]
167. Hews-Girard J, Guelcher C, Meldau J, McDonald E, Newall F. Principles and theory guiding development and delivery of patient education in disorders of thrombosis and hemostasis: reviewing the current literature. *Res Pract Thromb Haemost* 2017;1(2):162-171 [[FREE Full text](#)] [doi: [10.1002/rth2.12030](https://doi.org/10.1002/rth2.12030)] [Medline: [30046686](https://pubmed.ncbi.nlm.nih.gov/30046686/)]
168. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. *CoDesign* 2008;4(1):5-18. [doi: [10.1080/15710880701875068](https://doi.org/10.1080/15710880701875068)]
169. Tougas ME, Hayden JA, McGrath PJ, Huguette A, Rozario S. A systematic review exploring the social cognitive theory of self-regulation as a framework for chronic health condition interventions. *PLoS One* 2015;10(8):e0134977 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0134977](https://doi.org/10.1371/journal.pone.0134977)] [Medline: [26252889](https://pubmed.ncbi.nlm.nih.gov/26252889/)]
170. Orlando JF, Beard M, Kumar S. Systematic review of patient and caregivers' satisfaction with telehealth videoconferencing as a mode of service delivery in managing patients' health. *PLoS One* 2019;14(8):e0221848 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0221848](https://doi.org/10.1371/journal.pone.0221848)] [Medline: [31469865](https://pubmed.ncbi.nlm.nih.gov/31469865/)]
171. Miller CL, Mott K, Cousins M, Miller S, Johnson A, Lawson T, et al. Integrating consumer engagement in health and medical research - an Australian framework. *Health Res Policy Syst* 2017;15(1):9 [[FREE Full text](#)] [doi: [10.1186/s12961-017-0171-2](https://doi.org/10.1186/s12961-017-0171-2)] [Medline: [28187772](https://pubmed.ncbi.nlm.nih.gov/28187772/)]
172. Winpenny EM, Miani C, Pitchforth E, King S, Roland M. Improving the effectiveness and efficiency of outpatient services: a scoping review of interventions at the primary-secondary care interface. *J Health Serv Res Policy* 2017;22(1):53-64 [[FREE Full text](#)] [doi: [10.1177/1355819616648982](https://doi.org/10.1177/1355819616648982)] [Medline: [27165979](https://pubmed.ncbi.nlm.nih.gov/27165979/)]
173. Jessup R, Putrik P, Buchbinder R, Nezon J, Rischin K, Cyril S, et al. Identifying alternative models of healthcare service delivery to inform health system improvement: scoping review of systematic reviews. *BMJ Open* 2020;10(3):e036112 [[FREE Full text](#)] [doi: [10.1136/bmjopen-2019-036112](https://doi.org/10.1136/bmjopen-2019-036112)] [Medline: [32229525](https://pubmed.ncbi.nlm.nih.gov/32229525/)]

Abbreviations

CBT: cognitive behavioral therapy

COPD: chronic obstructive pulmonary disease

CVD: cardiovascular disease

HIC: high-income countries

ICT: information and communications technology

JBI: Joanna Briggs Institute

OECD: Organization for Economic Co-operation and Development

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

RCT: randomized controlled trial

SCT: social cognitive theory

Edited by G Eysenbach; submitted 15.12.20; peer-reviewed by E Thomas, W Zhang; comments to author 08.04.21; revised version received 15.04.21; accepted 18.11.21; published 09.03.22.

Please cite as:

*Beks H, King O, Clapham R, Alston L, Glenister K, McKinstry C, Quilliam C, Wellwood I, Williams C, Wong Shee A
Community Health Programs Delivered Through Information and Communications Technology in High-Income Countries: Scoping Review*

J Med Internet Res 2022;24(3):e26515

URL: <https://www.jmir.org/2022/3/e26515>

doi: [10.2196/26515](https://doi.org/10.2196/26515)

PMID: [35262498](https://pubmed.ncbi.nlm.nih.gov/35262498/)

©Hannah Beks, Olivia King, Renee Clapham, Laura Alston, Kristen Glenister, Carol McKinstry, Claire Quilliam, Ian Wellwood, Catherine Williams, Anna Wong Shee. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 09.03.2022. 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.

Review

Implementation of Cognitive Behavioral Therapy in e–Mental Health Apps: Literature Review

Kerstin Denecke¹, Dr rer nat; Nicole Schmid², MBA, LicPhilPs; Stephan Nüssli¹, Dr med

¹Institute for Medical Informatics, Bern University of Applied Sciences, Biel, Switzerland

²Suchtfachklinik Zurich, Zurich, Switzerland

Corresponding Author:

Kerstin Denecke, Dr rer nat

Institute for Medical Informatics

Bern University of Applied Sciences

Quellgasse 21

Biel, 2501

Switzerland

Phone: 41 32 321 67 94

Email: kerstin.denecke@bfh.ch

Abstract

Background: To address the matter of limited resources for treating individuals with mental disorders, e–mental health has gained interest in recent years. More specifically, mobile health (mHealth) apps have been suggested as electronic mental health interventions accompanied by cognitive behavioral therapy (CBT).

Objective: This study aims to identify the therapeutic aspects of CBT that have been implemented in existing mHealth apps and the technologies used. From these, we aim to derive research gaps that should be addressed in the future.

Methods: Three databases were screened for studies on mHealth apps in the context of mental disorders that implement techniques of CBT: PubMed, IEEE Xplore, and ACM Digital Library. The studies were independently selected by 2 reviewers, who then extracted data from the included studies. Data on CBT techniques and their technical implementation in mHealth apps were synthesized narratively.

Results: Of the 530 retrieved citations, 34 (6.4%) studies were included in this review. mHealth apps for CBT exploit two groups of technologies: technologies that implement CBT techniques for cognitive restructuring, behavioral activation, and problem solving (exposure is not yet realized in mHealth apps) and technologies that aim to increase user experience, adherence, and engagement. The synergy of these technologies enables patients to self-manage and self-monitor their mental state and access relevant information on their mental illness, which helps them cope with mental health problems and allows self-treatment.

Conclusions: There are CBT techniques that can be implemented in mHealth apps. Additional research is needed on the efficacy of the mHealth interventions and their side effects, including inequalities because of the digital divide, addictive internet behavior, lack of trust in mHealth, anonymity issues, risks and biases for user groups and social contexts, and ethical implications. Further research is also required to integrate and test psychological theories to improve the impact of mHealth and adherence to the e–mental health interventions.

(*J Med Internet Res* 2022;24(3):e27791) doi:[10.2196/27791](https://doi.org/10.2196/27791)

KEYWORDS

cognitive behavioral therapy; mHealth; e–mental health; chatbot; mobile phone

Introduction

Background

Mental disorders, including depressive and anxiety disorders, affect 29% of the global population in their lives [1]. Apart from the fact that mental disorders have an impact on people's quality of life, they are one of the most common causes of occupational

disability [2], resulting in high economic costs. The negative social aspects experienced by individuals with mental disorders include the inability to create and maintain lasting relationships and the stigmatization in society. These factors hamper individuals' capacity to act and lead a self-determined life as members of society, discourage them from seeking professional help, and possibly reinforce the characteristics of mental

disorders [3]. Mental disorders are usually treated using pharmacotherapy or psychotherapy [4]. However, there is a global shortage of mental health professionals as demand exceeds service provision. There are 9 psychiatrists per 100,000 people available in high-income countries [5], whereas there is 1 psychiatrist for every 10 million people in low-income countries [6]. In Europe, a comparative study between Finland and Spain—both with a similar prevalence of mental health disorders—showed a significant difference in the number of available staff resources. In Finland, for example, 13 psychologists were available per 100,000 inhabitants, whereas in Spain, only 2.9 psychologists were available per 100,000 inhabitants [7]. According to the World Health Organization (WHO), approximately 45% of people in high-income countries and 15% of people in low-income countries can access mental health services [8]. Leaving people with untreated mental disorders may increase suicide attempts and mortality [9]. Even if treated, approximately 98% of patients' change processes induced by therapy occur outside of therapy sessions in their daily lives. Therefore, there is a need to provide support and self-help between therapy sessions, which increases the availability of cognitive behavioral therapy (CBT) to larger populations.

To address the issue of limited resources for treating individuals with mental disorders, e-mental health has gained interest in recent years, particularly for behavior change using elements of CBT and self-help (eg, MoodGym [10]). e-mental health is defined as “mental health services and information delivered or enhanced through the Internet and related technologies.” [11]. Numerous studies have shown that e-mental health interventions are comparable in effectiveness to traditional face-to-face psychotherapy [12,13], thus providing a possible solution for people who do not have access to face-to-face therapy. e-Mental health enables users to learn more about their mental health condition through self-help services; it empowers them to strengthen their self-management and improve their health, sometimes including peer-to-peer support [14]. e-Mental health, realized as mobile health (mHealth) apps, aims to expand the availability and quality of mental health treatment. mHealth apps often ask users to enter data for reflection and awareness and provide relevant information depending on user input. Sometimes, they also collect data from wearables. The number of apps addressing mental health has rapidly increased in recent years [15,16]. Details on the technical implementation of mHealth apps are rarely described in scientific papers [17], although implementation is of utmost importance to enable patient agency and facilitate self-therapy practices. This study aims to investigate which CBT techniques are implemented by which technologies in mHealth apps and derive the research gaps.

CBT and mHealth Apps

CBT is an “active, problem-focused, and time-sensitive approach to treatment that aims to reduce emotional distress and increase adaptive behavior in patients with a host of mental health and

adjustment problems” [15,18]. There are four fundamental techniques of psychotherapy used in CBT: cognitive restructuring, behavioral activation, exposure, and problem solving [15,18]. In cognitive restructuring, therapists support patients in recognizing, evaluating, and modifying maladaptive or unhelpful thinking. Behavioral activation helps patients to actively re-engage in their lives. Exposure comprises systematic contact with a feared stimulus, whereas problem solving aims to help patients identify and implement solutions to their problems. We based our work on the implementation of CBT in mHealth apps on these 4 fundamental techniques of psychotherapy.

The efficacy of CBT has been demonstrated in multiple forms of psychopathology, including anxiety disorders, depression, and eating disorders [19]. Efficacy in the context of this study means that the effectiveness of an intervention can be demonstrated. mHealth apps provide options for practices, which were formerly elements of therapist-patient interaction; thus, this provides momentum for new routines and social forms [20] of coping with mental health problems.

In this study, we aim to assess which technologies are used in mHealth apps to implement the CBT technique. More specifically, we seek to answer the following research questions: for which mental illnesses have CBT-based apps been in use or tested, which CBT techniques are implemented in mHealth apps, which technologies are used to implement CBT techniques in mHealth apps, and which research gaps exist in mHealth apps for realizing CBT?

Methods

Overview

We answered our research questions using a literature search and review. We studied which of the 4 fundamental CBT techniques (cognitive restructuring, problem solving, behavioral activation, and exposure) have been implemented in mHealth apps. Furthermore, we investigated the technologies used for their implementation.

Search Strategy

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) criteria guided the conduct and reporting of our literature search [21]; for the PRISMA checklist, refer to [Multimedia Appendix 1](#). The search was conducted between June 6 and June 13, 2020, considering all articles published during the period of 2007 to 2020, as the first iPhone was launched in 2007, establishing the technological basis for mobile apps. A total of 3 databases were consulted to find the relevant papers. Papers included in PubMed were retrieved using the search string described in [Textbox 1](#). As we were also interested in the technical aspects of CBT in mHealth apps, we additionally searched the libraries of IEEE Xplore and ACM Digital Library ([Textbox 1](#)).

Textbox 1. Search strings used for database search.

Search strings used to search PubMed

- *Cognitive behavioral therapy AND mental health AND (telemedicine OR mobile health OR mhealth OR smartphone) NOT (internet delivered OR internet-delivered)*: 287 results with abstract

Search strings used to search IEEE Xplore

- *Cognitive behavioral therapy*: 65 results

Search strings used to search ACM Digital Library

- Query: (Abstract: *cognitive behavioral therapy AND mental health*) AND (Abstract: *mobile health OR mhealth OR smartphone*); Filter: (Article type: *Research Article*, Publication date: *[1/1/2007 TO *]*, ACM content: *DL*): 178 results

Inclusion and Exclusion Criteria

Articles were included in this review if they were dealing with CBT and mHealth apps, they were primary studies reporting results, and adults (aged >18 years) were the target population.

Articles were excluded if the target populations for interventions were military veterans, children, or adolescents. These target groups differ from the general public. Military veterans have a significantly higher prevalence of posttraumatic stress disorder than the average population [22] and often receive care in specialized institutions. On the other hand, digital resources for children or adolescents need to address specific cognitive and developmental issues [23] and cannot be directly compared with apps for adults. Therefore, we decided to exclude apps that explicitly targeted these groups from this review. Furthermore, all papers dealing only with web-based interventions and those describing only the study protocol without the final results were excluded.

Eligibility and Data Extraction

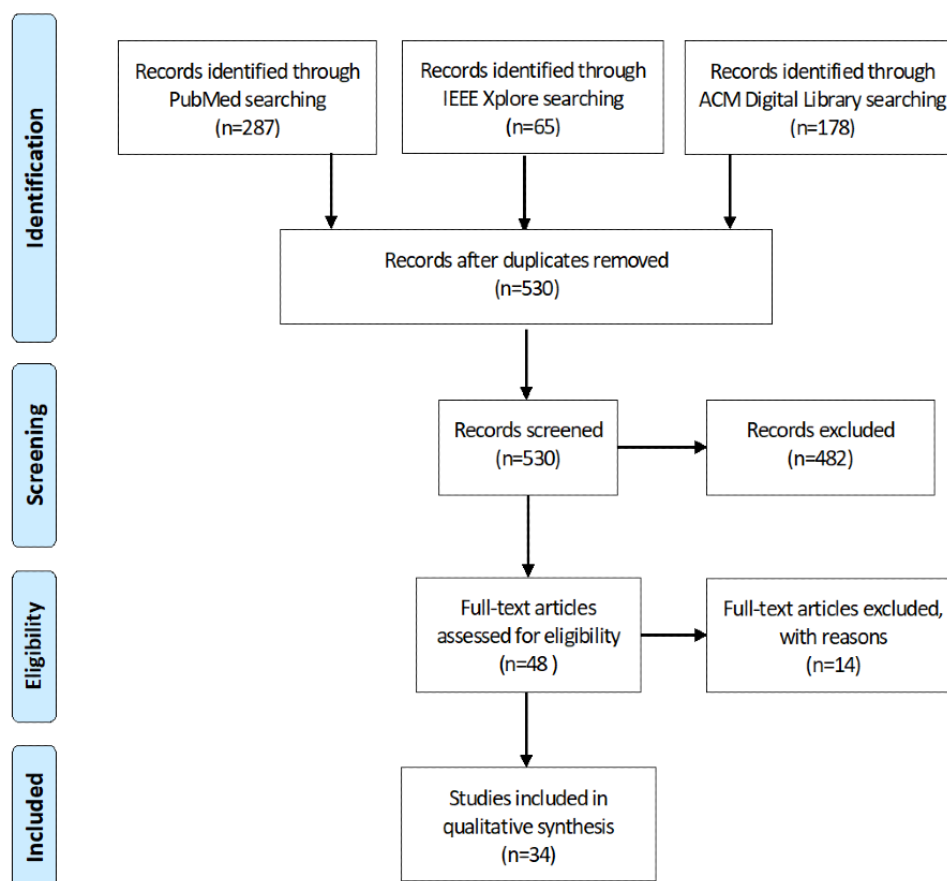
To assess the eligibility of the articles, all titles and abstracts were examined by 2 independent reviewers (KD and SN) in the first round. In the second round, the full texts of the selected articles were extracted and carefully analyzed to confirm their eligibility. Eligibility doubts were discussed until an agreement was reached. The selected articles were included in the qualitative synthesis.

Two reviewers (KD, NS) extracted data from the selected studies regarding CBT techniques, technologies implemented in the mHealth app, type of mHealth app, considered medical conditions, and outcome. With *CBT techniques*, we referred to the 4 fundamental techniques of psychotherapy applied within CBT (cognitive restructuring, behavioral activation, exposure, and problem solving [15]). With *technologies* in mHealth apps, we implied technical means used in the apps for realizing specific functionalities. More specifically, we assessed the provision of audio and video content, interactive elements (eg, communication facilities with humans or computer systems such as chatbots), social network technologies, gamification, and automatic analysis facilities (eg, for sentiment or emotion detection, recommendation, and text analysis). Data were abstracted into a spreadsheet standardized for this review. Finally, we derived the research gaps in e-mental health from the results.

Results

Sample

A total of 530 papers were retrieved by our search, as follows: 287 (54.2%) records in PubMed, 65 (12.3%) in IEEE Xplore, 178 (33.6%) in ACM Digital Library, and no duplicates. Of these 530 papers, 34 (6.4%) papers met the inclusion criteria and were, therefore, included in the qualitative synthesis (see the flowchart of the selection procedure in Figure 1). Papers were excluded during eligibility screening and data extraction if they described only mock-ups or no apps.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of the selection procedure.

Characteristics of Studies

Most papers (21/34, 62%) reported randomized controlled trials (RCTs) in which ≥ 1 mental health app was tested ([Multimedia Appendix 2](#)). The remaining papers used different study designs. Of the 34 papers, 4 (12%) only assessed usability through surveys. For RCTs, the number of involved participants varied between 30 and 1098, with an average of 333. The studies designed as surveys involved an average of 135 participants. We found the largest number of participants (n=3977) in an observational study [24]. This study did not recruit participants explicitly but analyzed the use protocols of the app under consideration. It remains unclear whether these protocols originated from people with diagnosed mental health problems.

Most studies considered usability [25,26], user satisfaction, acceptance [24], adherence [27], or engagement as outcome measures. RCTs mainly studied efficacy, adherence, or acceptability compared with a control group. In RCTs, e-mental health interventions were often compared against each other instead of comparing the app intervention with standard face-to-face psychotherapy. For example, the efficacy of a

web-based implementation of the Kokoro app was compared with that of an app-based intervention [28]. Among other things, outcome measures included the Patient Health Questionnaire–9 or Emotional Self-awareness Scale [29]. Of the 34 papers, 3 (9%) studied the Kokoro app for depression. In these papers, the effectiveness of the app-based intervention was reported; the use of the app reduced the symptoms of depression [28,30,31].

Mental Health Conditions Covered

Most studies (32/34, 94%) focused their research on a user group with a specific disorder. We classified the disorders according to the WHO International Classification of Diseases–10th revision (version 2019) coding system ([Table 1](#)). Most studies (16/34, 47%) used apps that offered support for persons with depression. Furthermore, there was a wide range of apps that targeted anxiety disorders; addiction problems; and psychiatric disorders such as schizophrenia, bipolar disorders, or suicidal tendencies. These disorders are often treated by CBT, which can be realized in an mHealth app in contrast to other types of therapies.

Table 1. Health conditions examined in the studies (N=34).

Health condition	WHO ^a ICD-10 ^b code	Total, n (%)
Drug addiction—cannabis	F12.1	1 (3)
Agoraphobia	F40.0	1 (3)
Borderline personality	F60.3	1 (3)
Depression	F32	16 (47)
Eating disorder	F50	2 (6)
Insomnia	G47	1 (3)
Psychosis	F29	2 (6)
Schizophrenia, schizoaffective, or bipolar 1	F20, F25, and F31	1 (3)
Smoking	Z72.0	3 (9)
Stress	F99	3 (9)
Suicidality	R45.8	1 (3)
Unknown	N/A ^c	2 (6)

^aWHO: World Health Organization.

^bICD-10: International Classification of Diseases–10th revision.

^cN/A: not applicable.

CBT Techniques in mHealth Apps

Apps may have integrated techniques to address different purposes. We classified the main purposes of the apps as informational, coach, or therapy. An app was considered informational when it mainly provided information to the user, for example, on mental health. Apps were classified as coaches when suggestions were tailored to the user and their specific mental health condition or when support in managing a mental illness was delivered. As therapeutic apps, we grouped apps that delivered CBT with the aim of creating a therapeutic setting.

We could not identify any apps that were only informational. Most apps provided information but had additional functionalities that led to their classification as coach or therapy. Most (20/34, 59%) apps could be considered as coaches in mental health. Of the 34 apps, 13 (38%) were classified as therapy; for 1 (3%) app, this classification was not applicable. We concluded that CBT-based mHealth apps go beyond patient education. They support and extend their self-help capabilities.

We identified in the included papers 3 of the 4 fundamental techniques of psychotherapy applied in CBT (cognitive restructuring, behavioral activation, exposure, and problem solving [15]; [Multimedia Appendix 3](#) [24-27,29,30,32-50]). mHealth apps provided toolkits for cognitive restructuring, which included diary-keeping functionalities and support in changing thoughts or tensions. Behavioral activation was realized by providing information on mental health conditions by tracking activities or setting goals. Assigning homework to patients, which is the best practice as a problem-solving technique [15], was realized in many apps (20/34, 59%) by providing exercises and activities as strategies to cope with mental health problems. Direct communication with a therapist, which is not a CBT technique itself but helps in standard setting for delivering CBT, was enabled by 6% (2/34) of the apps included in the review.

Technologies in mHealth Apps for Implementing CBT

When developing mHealth apps for CBT, it is important to know which technologies are useful and efficient for implementing different CBT techniques. Interactive elements were included in 35% (12/34) of apps to realize behavioral activation and cognitive restructuring. Among other things, we could identify the following interactive elements: automatic question answering functionalities, message exchange with the treatment team, a virtual character that provided information or explanations, personalization, and persuasion methods ([Multimedia Appendix 4](#) [24-28,31-37,40-46,50-54]). Of the 34 apps, 3 (8%) provided a conversational agent or chatbot with which the user could communicate. Chatbot technology was used to support cognitive restructuring. Methods for automatic analysis of free textual input (natural language processing) and machine learning methods (eg, text classification or clustering) were used to understand user input or personalize recommendations according to user input. However, these methods are still rarely used (3/34, 9% papers). Audio and video were used to provide information or demonstrate exercises such as meditation exercises (ie, support behavioral activation and problem solving). Few apps calculated scores such as sentiment or emotion scores that were shown on a timeline. Sentiment or emotion scores quantify the sentiment or emotion of a user (eg, positive and negative sentiment or the strength of an emotion such as sadness). The following two technologies were integrated into some apps that aimed to improve the user experience and adherence: gamification and social networks. Social aspects were integrated, enabling the user to connect with other users. This social community aspect was interesting in the context of depression but also in other mental diseases, where people often experience loneliness and a lack of social contact. Gamification and social networks integrated into these apps did not aim to deliver CBT but, more importantly, aimed to increase the adherence and attractiveness of using the apps

[32,51]. Gamification such as collecting jigsaws was used in 24% (8/34) of apps, and audio or video content was provided by 21% (7/34) of apps. Approximately 15% (5/34) of apps enabled connections with social networks and communities.

Discussion

Principal Findings

Overview

The main finding of this review is that mHealth apps for CBT exploit two groups of technologies: (1) technologies that implement concrete CBT techniques for cognitive restructuring, behavioral activation, and problem solving and (2) technologies that aim to increase user experience, adherence, and engagement. The latter tries to address the current challenge in delivering CBT, which is insufficient adherence to CBT techniques [55]. A CBT technique that we did not find implemented was *exposure*. There is a broad range of technologies used in mHealth apps to deliver CBT; however, no technology was used in all apps. However, a trend toward the use of interactive elements, gamification technologies, and technologies supporting social activities could be recognized. No app was purely informational; however, most could be classified as coaches and a few as therapeutic. There have been attempts to deliver therapy through such apps [33-35,52]. Most apps target patients with disorders that are often treated with CBT, such as depression; anxiety disorders; addiction problems; and psychiatric disorders such as schizophrenia, bipolar disorders, or suicidal tendencies.

Mental Illnesses for Which CBT-Based Apps Have Been Used or Tested

Depression was reported most often in the assessed studies. Depressive disorders are widespread in all countries and comorbid with other mental diseases. According to the WHO, >264 million people of all ages experience depression worldwide. Thus, the target user group is huge. Depression is a leading cause of disability worldwide and a major contributor to the overall global burden of disease. There seems to be a burden of being able to deliver support to those people, and mental health apps are obviously comprehensively tested to fill this gap. Although our review did not deliver much evidence on the efficacy of mHealth apps in the mental health context, Khademian et al [56] found that mHealth apps that provide behavior change strategies, such as CBT and techniques for behavioral activation, have significant effects on depression, anxiety, and stress.

Psychiatric diseases such as psychosis or schizophrenia were also targeted. A challenge with psychiatric diseases is that critical situations can occur where professional reactions are essential (eg, to prevent a suicide attempt [57]). mHealth apps that deliver predefined content cannot react individually to various situations or specific user needs. There is a need for mHealth apps without undesired side effects; that is, those able to respond appropriately in situations of crisis [17].

CBT Techniques Implemented in mHealth Apps

We found that some methods used in standard CBT were implemented in mHealth apps. Sometimes, it was simply a digitization of the existing technique; however, there were cases where mHealth apps offered some benefits compared with the traditional therapy setting. Thought records or coping cards are methods used within standard CBT to achieve cognitive restructuring [15]. These methods are often provided in mHealth apps, for example, by supporting the keeping of diaries.

Scheduling and monitoring activities are the central components of behavioral activation in CBT [58]. Behavioral activation recommends planning activities in the evening before or in the early morning. As therapy sessions are spread across weeks and planning tasks are conducted with the therapist, day-to-day planning is not practiced in traditional therapy settings. With the use of mHealth technologies, daily planning and engagement can be supported, and personalized recommendations can be made [36]. Furthermore, increasing patients' confidence in their ability to cope with stress and adversity and their overall coping skills is a major factor in mHealth apps contributing to the effects on mental health [29].

Information provision as a means of behavioral activation is another important aspect realized in many apps. This task includes teaching patients the basics of their mental health condition and CBT techniques. Morriss et al [59,60] have shown that knowledge and understanding of the medical condition are effective in supporting the everyday coping of patients with mental illness and can foster compliance of patients to the health intervention. Furthermore, patients become more competent in making decisions related to their health through information provision [61]. As providing information is a repetitive task for health care providers, an app providing information and teaching patients on the basics of mental health could save time for health care providers. Furthermore, the patient can read the information several times. However, studies show that e-mental health apps often contain incorrect information [62] and are not necessarily evidence based.

Assigning homework to patients is the best practice as a problem solving technique [15]. This aspect of CBT, which is available in many apps, is realized by providing exercises and activities. Exposure (imaginal exposure and interoceptive exposure [15]) is a CBT technique that was not represented at all in mental health apps included in our review. However, recent research has exploited virtual reality for realizing exposure apps; for example, to reduce the fear of heights [63]. It is obvious that there is still potential for developing and testing mHealth apps that target exposure techniques.

In conclusion, mental health apps that provide CBT support the performing of repetitive tasks such as keeping a diary or exercising, which are essential tasks and techniques within CBT for cognitive restructuring, behavioral activation, and problem solving [15]. Although these tasks must be realized by patients as part of their standard therapy, even without mHealth support, their digitization in mHealth apps can increase adherence, user engagement, and retention or facilitate learning [64]. A mobile app is available 24 hours a day. Thus, monitoring and tracking can be performed at any time and at any place. Records in the

app cannot get lost as paper-based records do. Reminders can be sent to the user on a regular basis to ensure that records are made or activities come into the user's mind [37].

Technologies Used to Implement CBT in mHealth Apps

Most apps often present information using videos or audio, or they collect data on activities from a user for behavioral activation. Social aspects are integrated, enabling the user to connect with other users. This social community aspect is interesting not only in the context of depression but also other mental diseases, where people often experience loneliness and a lack of social contact. Gamification and social networks integrated in these apps do not aim to deliver CBT but, more importantly, aim to increase the adherence and attractiveness of using these apps [37]. Artificial intelligence, including natural language processing, language understanding, and chatbot technology, has been rarely used in available apps. A reason might be that unforeseen errors or reactions can occur when the system misinterprets the user input, which might be avoided [65]. Other studies have shown an increased interest in mental health chatbots [66] or that chatbots are examples of the next generation of mental health [17]. However, this was not reflected in our review. Another reason for the limited inclusion of artificial intelligence in mental health apps may be related to medical device regulation, which requires traceability and increases the demand on the development process.

Research Gaps That Exist in mHealth Apps for Realizing CBT

After scoping the landscape of mHealth apps that implement CBT, we summarized the open research issues. There is a need for the following: RCTs studying the efficacy of the single technologies implemented in mHealth apps for realizing CBT techniques; studies on the impact on patients' agency, including trust and overreliance; consideration of psychological theories during mHealth implementation to increase impact and adherence; and support in recommending or selecting CBT apps as health interventions.

Although more than half of the analyzed studies reported on RCTs, which is the state-of-the-art study design for proving the efficacy of medical interventions [67], these trials often did not assess the efficacy of the mHealth intervention. Similar results have been reported by Bauer et al [68]. For chatbots in clinical psychology, Bendig et al [17] noted that studies on mental health chatbots mainly assessed feasibility and acceptance. Studies are needed to assess which technology is well-suited for implementing a particular CBT technique, as well as for which mental disorders are mHealth apps efficient. This will help derive the best practices for implementing CBT techniques in mHealth. In this context, it is also important to clarify the role of mHealth apps; for example, whether they are intended to accompany the therapy or fill the treatment gap until a therapy can be started.

We recognize that the impact of mHealth apps on patients' agency is not explicitly considered in these studies. Agency is generally viewed as the "capacity to act, produce and anticipate a desired outcome within a particular context" [69]. In our context, this means, among other things, that through mental

health apps, patients would begin to manage themselves as participants with agency exercised on the basis of patient autonomy [70].

From a technical perspective, we want to achieve good adherence and frequent use of apps and support this by integrating features such as gamification to encourage frequent use. However, there are also concerns about overreliance on mental health apps, particularly for people who already have trouble with addictive web-based behavior [71]. The use of smartphone apps might be problematic, given specific mental conditions such as anxiety or depression [72]. Studies that assess this risk of overreliance in long-term use of mHealth apps, as well as assessments related to trust in mHealth from perspectives of both patients and therapists, are missing. There are indications that users think they have even more trust in an mHealth app when reporting personal issues than in their physician [73]. Anonymity helps overcome thoughts of stigmatization or shame when describing personal issues [74].

We are also missing implementations of psychological theories in CBT mental health apps to increase the impact of and adherence to these technological means. Peters et al [75] suggested a framework grounded in psychological research that could help developers of mHealth apps develop apps that increase motivation and engagement and, in this way, also have an impact on patients' well-being.

Recommendation of Apps

Our literature search showed that the underlying clinical evidence, technical issues, and implementation details are rarely described in the published research on mental health apps. Without information on the correctness of the underlying evidence base and without confirmation of efficacy, it is challenging for patients to identify the mental health apps that are suited for them.

Therapists and patients rely on lists of *top mental health apps* for their decision-making process, which are not helpful unless they list their ranking and selection criteria for creating the list. We checked for 21 apps listed on the mental health app ranking lists to determine whether there were publications available. A scientific paper or results from a clinical trial were available for only 29% (6/21) of the apps. This makes it difficult for users and health professionals to identify high-quality, evidence-based apps. It must be ensured that the underlying principles and delivered contents are evidence based [17,76] to avoid harm to users, as well as undesired consequences for the users. The number of downloads from an app store is clearly not an indicator of high-quality mHealth apps. On the other hand, we found studies on 28 apps, of which only 16 (57%) were implemented for the established Android and iOS operating systems. Of these 16 apps, only 12 (75%) were available in the stores. This means that apps that have been thoroughly investigated are not available to a wide range of users.

We conclude that there is a need to enable patients in informed decision-making when selecting an mHealth app for CBT [77]. We suggest a minimal data set of information on a mental health app that includes the following items: underlying clinical evidence and CBT techniques that are integrated; information

on how CBT techniques are implemented in the app, that is, which technologies are used and for which purpose; efficacy from an RCT, if available; intended application areas of the app; possible contraindications; information on data storage, data security, and privacy; and the integrated third-party tools.

Standardized evaluation metrics should be developed to ensure that only high-quality apps are recommended by therapists. Attempts have been made to harmonize technical metrics for evaluating health chatbots [78,79]. Existing evaluation frameworks for mHealth apps target usability [76]. Although good usability is essential to make sure that users can interact with the app, it is highly relevant that the app really positively affects patients and does not harm them.

Strengths and Limitations of Our Presented Research Work

Strengths

This literature review was conducted to identify the technologies used in mHealth apps for delivering CBT and identify the research gaps. To the best of our knowledge, this study is among the first to summarize these aspects and provide an overview. This study helps in identifying the opportunities and limitations of mHealth in supporting CBT. It also shows the potential and limitations of e-mental health with respect to the self-help capabilities of patients.

As 2 reviewers independently selected the articles and extracted the data, selection bias in this review was reduced. There is a review of Stawarz et al [80] that studies user experience of CBT-enabling apps [80]. In contrast to their work, we are more interested in the technical implementation of CBT in mHealth apps and the research gaps and not only in the features provided by the apps.

Limitations

The search in this review was restricted to English and German articles. Accordingly, it is likely that this review missed some publications. We also missed apps for which no publications were available. In addition, we restricted the search to papers that included the terms *cognitive behavior therapy* or *mHealth* or the respective Medical Subject Headings terms. We might have missed relevant papers that did not explicitly use these

terms but dealt with a relevant topic. Several terms can be related to mHealth; however, these are too broad to be reviewed in a reasonable amount of time.

Conclusions

In this paper, we studied the therapeutic aspects of CBT that are implemented in mHealth apps and the technologies by which they are integrated into these apps. We conclude that some CBT techniques (behavioral activation and cognitive restructuring) can be well-realized in a mental health app, whereas others are more difficult to implement (problem solving and exposure). mHealth apps for CBT can support patients through additional self-help and self-management tools that support specific aspects of treatment. mHealth apps facilitate the self-competence and self-management of patients in coping with mental health problems. Interactive elements, gamification, and integration of social networks are technologies that increase user engagement and adherence. In this way, patients are provided with additional action and interaction capabilities. As they cover relevant aspects of CBT treatment, mHealth apps can, in principle, alleviate the issue of a global shortage of mental health human resources, as they are usually available to millions of users anytime and anywhere, presuming the users have mobile devices and internet access. This, in turn, has the potential to improve the availability and quality of mental health care at reduced expenses. However, these apps must be evidence based and reliable to be effective and avoid harm to patients. Further research is needed on the side effects of mHealth self-help treatments, such as new inequalities because of the digital divide, addictive internet behavior, trust in mHealth, anonymity issues, risk and range of self-help treatments, bias toward user groups and social contexts, and ethical implications.

The future role of app-based e-mental health still has to be clarified. Future research should assess integrated care models where all stakeholders can benefit: the health care provider by avoiding repetitive tasks and the patient by receiving support even beyond therapy sessions. In addition, economic value is relevant for bringing evidence-based apps to users and supporting clinical care. For obsessive-compulsive disorders, it has already been shown that computerized CBT delivered as low-intensity interventions is cost-effective [81].

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.

[DOCX File, 18 KB - [jmir_v24i3e27791_app1.docx](#)]

Multimedia Appendix 2

Study designs reported in the papers.

[DOCX File, 13 KB - [jmir_v24i3e27791_app2.docx](#)]

Multimedia Appendix 3

Cognitive behavioral therapy tools and methods realized in the reported apps.

[DOCX File , 14 KB - [jmir_v24i3e27791_app3.docx](#)]

Multimedia Appendix 4

Technologies used in the apps.

[DOCX File , 13 KB - [jmir_v24i3e27791_app4.docx](#)]

References

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. *Int J Epidemiol* 2014;43(2):476-493 [FREE Full text] [doi: [10.1093/ije/dyu038](#)] [Medline: [24648481](#)]
2. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. *PLoS One* 2015;10(2):e0116820 [FREE Full text] [doi: [10.1371/journal.pone.0116820](#)] [Medline: [25658103](#)]
3. Angermeyer MC, Matschinger H. The stigma of mental illness: effects of labelling on public attitudes towards people with mental disorder. *Acta Psychiatr Scand* 2003;108(4):304-309. [doi: [10.1034/j.1600-0447.2003.00150.x](#)] [Medline: [12956832](#)]
4. Cuijpers P, Sijbrandij M, Koole SL, Andersson G, Beekman AT, Reynolds 3rd CF. The efficacy of psychotherapy and pharmacotherapy in treating depressive and anxiety disorders: a meta-analysis of direct comparisons. *World Psychiatry* 2013;12(2):137-148 [FREE Full text] [doi: [10.1002/wps.20038](#)] [Medline: [23737423](#)]
5. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* 2012;380(9859):2197-2223. [doi: [10.1016/S0140-6736\(12\)61689-4](#)] [Medline: [23245608](#)]
6. Oladeji BD, Gureje O. Brain drain: a challenge to global mental health. *BJPsych Int* 2016;13(3):61-63 [FREE Full text] [doi: [10.1192/s2056474000001240](#)] [Medline: [29093905](#)]
7. Sadeniemi M, Almeda N, Salinas-Pérez JA, Gutiérrez-Colosía MR, García-Alonso C, Ala-Nikkola T, et al. A comparison of mental health care systems in northern and southern Europe: a service mapping study. *Int J Environ Res Public Health* 2018;15(6):1133 [FREE Full text] [doi: [10.3390/ijerph15061133](#)] [Medline: [29857556](#)]
8. Anthes E. Mental health: there's an app for that. *Nature* 2016;532(7597):20-23. [doi: [10.1038/532020a](#)] [Medline: [27078548](#)]
9. Hester RD. Lack of access to mental health services contributing to the high suicide rates among veterans. *Int J Ment Health Syst* 2017;11:47 [FREE Full text] [doi: [10.1186/s13033-017-0154-2](#)] [Medline: [28828036](#)]
10. MoodGym. URL: <https://moodgym.com.au>, [accessed 2021-02-04]
11. Christensen H, Griffiths KM, Evans K. e-mental health in Australia: implications of the internet and related technologies for policy. Canberra, Australia: Commonwealth Department of Health and Ageing; 2002.
12. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. *World Psychiatry* 2014;13(3):288-295 [FREE Full text] [doi: [10.1002/wps.20151](#)] [Medline: [25273302](#)]
13. 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;47(1):1-18. [doi: [10.1080/16506073.2017.1401115](#)] [Medline: [29215315](#)]
14. Moock J. Support from the internet for individuals with mental disorders: advantages and disadvantages of e-mental health service delivery. *Front Public Health* 2014;2:65 [FREE Full text] [doi: [10.3389/fpubh.2014.00065](#)] [Medline: [24967221](#)]
15. Wenzel A. Basic strategies of cognitive behavioral therapy. *Psychiatr Clin North Am* 2017;40(4):597-609 [FREE Full text] [doi: [10.1016/j.psc.2017.07.001](#)] [Medline: [29080588](#)]
16. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. *Mhealth* 2018;4:6 [FREE Full text] [doi: [10.21037/mhealth.2018.03.02](#)] [Medline: [29682510](#)]
17. Bendig E, Erb B, Schulze-Thuesing L, Baumeister H. The next generation: chatbots in clinical psychology and psychotherapy to foster mental health – a scoping review. *Verhaltenstherapie* 2019;1-13 [FREE Full text] [doi: [10.1159/000501812](#)]
18. Huppert JD. The building blocks of treatment in cognitive-behavioral therapy. *Isr J Psychiatry Relat Sci* 2009;46(4):245-250 [FREE Full text] [Medline: [20635770](#)]
19. Butler AC, Chapman JE, Forman EM, Beck AT. The empirical status of cognitive-behavioral therapy: a review of meta-analyses. *Clin Psychol Rev* 2006;26(1):17-31. [doi: [10.1016/j.cpr.2005.07.003](#)] [Medline: [16199119](#)]
20. Reckwitz A. Toward a theory of social practices. *Eur J Soc Theory* 2002;5(2):243-263 [FREE Full text] [doi: [10.1177/1368431022225432](#)]
21. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62(10):1006-1012. [doi: [10.1016/j.jclinepi.2009.06.005](#)] [Medline: [19631508](#)]
22. Stefanovics EA, Potenza MN, Pietrzak RH. PTSD and obesity in U.S. military veterans: prevalence, health burden, and suicidality. *Psychiatry Res* 2020;291:113242. [doi: [10.1016/j.psychres.2020.113242](#)] [Medline: [32763532](#)]

23. Martens M. Issues of access and usability in designing digital resources for children. *Libr Inf Sci Res* 2012;34(3):159-168 [FREE Full text] [doi: [10.1016/j.lisr.2011.12.003](https://doi.org/10.1016/j.lisr.2011.12.003)]
24. Plaza García I, Sánchez CM, Espílez Á, García-Magariño I, Guillén GA, García-Campayo J. Development and initial evaluation of a mobile application to help with mindfulness training and practice. *Int J Med Inform* 2017;105:59-67. [doi: [10.1016/j.ijmedinf.2017.05.018](https://doi.org/10.1016/j.ijmedinf.2017.05.018)] [Medline: [28750912](https://pubmed.ncbi.nlm.nih.gov/28750912/)]
25. Denecke K, Vaaheesan S, Arulnathan A. A mental health chatbot for regulating emotions (SERMO) - concept and usability test. *IEEE Trans Emerg Topics Comput* 2021;9(3):1170-1182 [FREE Full text] [doi: [10.1109/tetc.2020.2974478](https://doi.org/10.1109/tetc.2020.2974478)]
26. Purkayastha S, Addepally SA, Bucher S. Engagement and usability of a cognitive behavioral therapy mobile app compared with web-based cognitive behavioral therapy among college students: randomized heuristic trial. *JMIR Hum Factors* 2020;7(1):e14146 [FREE Full text] [doi: [10.2196/14146](https://doi.org/10.2196/14146)] [Medline: [32012043](https://pubmed.ncbi.nlm.nih.gov/32012043/)]
27. Martínez-Miranda J, Martínez A, Ramos R, Aguilar H, Jiménez L, Arias H, et al. Assessment of users' acceptability of a mobile-based embodied conversational agent for the prevention and detection of suicidal behaviour. *J Med Syst* 2019;43(8):246. [doi: [10.1007/s10916-019-1387-1](https://doi.org/10.1007/s10916-019-1387-1)] [Medline: [31240494](https://pubmed.ncbi.nlm.nih.gov/31240494/)]
28. Furukawa TA, Imai H, Horikoshi M, Shimodera S, Hiroe T, Funayama T, FLATT Investigators. Behavioral activation: is it the expectation or achievement, of mastery or pleasure that contributes to improvement in depression? *J Affect Disord* 2018;238:336-341. [doi: [10.1016/j.jad.2018.05.067](https://doi.org/10.1016/j.jad.2018.05.067)] [Medline: [29906719](https://pubmed.ncbi.nlm.nih.gov/29906719/)]
29. 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;109:75-83. [doi: [10.1016/j.brat.2018.08.003](https://doi.org/10.1016/j.brat.2018.08.003)] [Medline: [30125790](https://pubmed.ncbi.nlm.nih.gov/30125790/)]
30. Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: randomized controlled trial. *J Med Internet Res* 2017;19(11):e373 [FREE Full text] [doi: [10.2196/jmir.8602](https://doi.org/10.2196/jmir.8602)] [Medline: [29101095](https://pubmed.ncbi.nlm.nih.gov/29101095/)]
31. Furukawa TA, Horikoshi M, Fujita H, Tsujino N, Jinnin R, Kako Y, et al. Cognitive and behavioral skills exercises completed by patients with major depression during smartphone cognitive behavioral therapy: secondary analysis of a randomized controlled trial. *JMIR Ment Health* 2018;5(1):e4 [FREE Full text] [doi: [10.2196/mental.9092](https://doi.org/10.2196/mental.9092)] [Medline: [29326098](https://pubmed.ncbi.nlm.nih.gov/29326098/)]
32. Monney G, Penzenstadler L, Dupraz O, Etter JF, Khazaal Y. mHealth app for cannabis users: satisfaction and perceived usefulness. *Front Psychiatry* 2015;6:120 [FREE Full text] [doi: [10.3389/fpsy.2015.00120](https://doi.org/10.3389/fpsy.2015.00120)] [Medline: [26379561](https://pubmed.ncbi.nlm.nih.gov/26379561/)]
33. Christoforou M, Sáez Fonseca JA, Tsakanikos E. Two novel cognitive behavioral therapy-based mobile apps for agoraphobia: randomized controlled trial. *J Med Internet Res* 2017;19(11):e398 [FREE Full text] [doi: [10.2196/jmir.7747](https://doi.org/10.2196/jmir.7747)] [Medline: [29175809](https://pubmed.ncbi.nlm.nih.gov/29175809/)]
34. Depp CA, Perivoliotis D, Holden J, Dorr J, Granholm EL. Single-session mobile-augmented intervention in serious mental illness: a three-arm randomized controlled trial. *Schizophr Bull* 2019;45(4):752-762 [FREE Full text] [doi: [10.1093/schbul/sby135](https://doi.org/10.1093/schbul/sby135)] [Medline: [30281086](https://pubmed.ncbi.nlm.nih.gov/30281086/)]
35. Prada P, Zamberg I, Bouillault G, Jimenez N, Zimmermann J, Hasler R, et al. EMOTEO: a smartphone application for monitoring and reducing aversive tension in borderline personality disorder patients, a pilot study. *Perspect Psychiatr Care* 2017;53(4):289-298. [doi: [10.1111/ppc.12178](https://doi.org/10.1111/ppc.12178)] [Medline: [27439663](https://pubmed.ncbi.nlm.nih.gov/27439663/)]
36. Rohani DA, Lopategui AQ, Tuxen N, Faurholt-Jepsen M, Kessing LV, Bardram JE. MUBS: a personalized recommender system for behavioral activation in mental health. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020 Presented at: CHI '20; April 25-30, 2020; Honolulu, HI p. 1-13. [doi: [10.1145/3313831.3376879](https://doi.org/10.1145/3313831.3376879)]
37. Chen YX, Chen HC, Chen LX, Hu JW, Shie CK, Lin YS, et al. Enhancing adherence to cognitive behavioral therapy for insomnia through machine and social persuasion. In: *IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*. 2013 Presented at: CPSCom & GreenCom '13; August 20-23, 2013; Beijing, China p. 750-757 URL: <https://doi.org/10.1109/greencom-ithings-cpscom.2013.135> [doi: [10.1109/greencom-ithings-cpscom.2013.135](https://doi.org/10.1109/greencom-ithings-cpscom.2013.135)]
38. O'Connor M, Whelan R, Bricker J, McHugh L. Randomized controlled trial of a smartphone application as an adjunct to acceptance and commitment therapy for smoking cessation. *Behav Ther* 2020;51(1):162-177. [doi: [10.1016/j.beth.2019.06.003](https://doi.org/10.1016/j.beth.2019.06.003)] [Medline: [32005334](https://pubmed.ncbi.nlm.nih.gov/32005334/)]
39. Nobis S, Lehr D, Ebert DD, Berking M, Heber E, Baumeister H, et al. Efficacy and cost-effectiveness of a web-based intervention with mobile phone support to treat depressive symptoms in adults with diabetes mellitus type 1 and type 2: design of a randomised controlled trial. *BMC Psychiatry* 2013;13:306 [FREE Full text] [doi: [10.1186/1471-244X-13-306](https://doi.org/10.1186/1471-244X-13-306)] [Medline: [24238346](https://pubmed.ncbi.nlm.nih.gov/24238346/)]
40. Scherr S, Goering M. Is a self-monitoring app for depression a good place for additional mental health information? Ecological momentary assessment of mental help information seeking among smartphone users. *Health Commun* 2020;35(8):1004-1012. [doi: [10.1080/10410236.2019.1606135](https://doi.org/10.1080/10410236.2019.1606135)] [Medline: [31025888](https://pubmed.ncbi.nlm.nih.gov/31025888/)]
41. Rohani D, Tuxen N, Lopategui AQ, Faurholt-Jepsen M, Kessing LV, Bardram JE. Personalizing mental health: a feasibility study of a mobile behavioral activation tool for depressed patients. In: *Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 2019 Presented at: PervasiveHealth '19; May 20-23, 2019; Trento, Italy p. 282-291. [doi: [10.1145/3329189.3329214](https://doi.org/10.1145/3329189.3329214)]

42. Bucci S, Barrowclough C, Ainsworth J, Machin M, Morris R, Berry K, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. *Schizophr Bull* 2018;44(5):1070-1080 [FREE Full text] [doi: [10.1093/schbul/sby032](https://doi.org/10.1093/schbul/sby032)] [Medline: [29566206](https://pubmed.ncbi.nlm.nih.gov/29566206/)]
43. Anastasiadou D, Folkvord F, Brugnera A, Cañas Vinader L, SerranoTroncoso E, Carretero Jardí C, et al. An mHealth intervention for the treatment of patients with an eating disorder: a multicenter randomized controlled trial. *Int J Eat Disord* 2020;53(7):1120-1131. [doi: [10.1002/eat.23286](https://doi.org/10.1002/eat.23286)] [Medline: [32383503](https://pubmed.ncbi.nlm.nih.gov/32383503/)]
44. Tudor-Sfetea C, Rabee R, Najim M, Amin N, Chadha M, Jain M, et al. Evaluation of two mobile health apps in the context of smoking cessation: qualitative study of cognitive behavioral therapy (CBT) versus non-CBT-based digital solutions. *JMIR Mhealth Uhealth* 2018;6(4):e98 [FREE Full text] [doi: [10.2196/mhealth.9405](https://doi.org/10.2196/mhealth.9405)] [Medline: [29669708](https://pubmed.ncbi.nlm.nih.gov/29669708/)]
45. Clarke J, Proudfoot J, Birch MR, Whitton AE, Parker G, Manicavasagar V, et al. Effects of mental health self-efficacy on outcomes of a mobile phone and web intervention for mild-to-moderate depression, anxiety and stress: secondary analysis of a randomised controlled trial. *BMC Psychiatry* 2014;14:272 [FREE Full text] [doi: [10.1186/s12888-014-0272-1](https://doi.org/10.1186/s12888-014-0272-1)] [Medline: [25252853](https://pubmed.ncbi.nlm.nih.gov/25252853/)]
46. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. *J Med Internet Res* 2019;21(6):e12556 [FREE Full text] [doi: [10.2196/12556](https://doi.org/10.2196/12556)] [Medline: [31199319](https://pubmed.ncbi.nlm.nih.gov/31199319/)]
47. Bardram JE, Rohani DA, Tuxen N, Faurholt-Jepsen M, Kessing LV. Supporting smartphone-based behavioral activation: a simulation study. In: *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers*. 2017 Presented at: UbiComp '17; September 11-15, 2017; Maui, HI p. 830-843. [doi: [10.1145/3123024.3125617](https://doi.org/10.1145/3123024.3125617)]
48. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. *PLoS One* 2018;13(12):e0209482 [FREE Full text] [doi: [10.1371/journal.pone.0209482](https://doi.org/10.1371/journal.pone.0209482)] [Medline: [30596696](https://pubmed.ncbi.nlm.nih.gov/30596696/)]
49. 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;212(4):199-206. [doi: [10.1192/bjp.2018.10](https://doi.org/10.1192/bjp.2018.10)] [Medline: [30071909](https://pubmed.ncbi.nlm.nih.gov/30071909/)]
50. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. *JMIR Mhealth Uhealth* 2019;7(6):e14273 [FREE Full text] [doi: [10.2196/14273](https://doi.org/10.2196/14273)] [Medline: [31237569](https://pubmed.ncbi.nlm.nih.gov/31237569/)]
51. Tregarthen J, Paik Kim J, Sadeh-Sharvit S, Neri E, Welch H, Lock J. Comparing a tailored self-help mobile app with a standard self-monitoring app for the treatment of eating disorder symptoms: randomized controlled trial. *JMIR Ment Health* 2019;6(11):e14972 [FREE Full text] [doi: [10.2196/14972](https://doi.org/10.2196/14972)] [Medline: [31750837](https://pubmed.ncbi.nlm.nih.gov/31750837/)]
52. Kim SW, Lee GY, Yu HY, Jung EI, Lee JY, Kim SY, et al. Development and feasibility of smartphone application for cognitive-behavioural case management of individuals with early psychosis. *Early Interv Psychiatry* 2018;12(6):1087-1093. [doi: [10.1111/eip.12418](https://doi.org/10.1111/eip.12418)] [Medline: [28516480](https://pubmed.ncbi.nlm.nih.gov/28516480/)]
53. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. *J Med Internet Res* 2016;18(12):e330 [FREE Full text] [doi: [10.2196/jmir.6482](https://doi.org/10.2196/jmir.6482)] [Medline: [27998876](https://pubmed.ncbi.nlm.nih.gov/27998876/)]
54. Cheng P, Kalmbach DA, Tallent G, Joseph CL, Espie CA, Drake CL. Depression prevention via digital cognitive behavioral therapy for insomnia: a randomized controlled trial. *Sleep* 2019;42(10):zsz150 [FREE Full text] [doi: [10.1093/sleep/zsz150](https://doi.org/10.1093/sleep/zsz150)] [Medline: [31535688](https://pubmed.ncbi.nlm.nih.gov/31535688/)]
55. Edmunds JM, Brodman DM, Ringle VA, Read KL, Kendall PC, Beidas RS. Examining adherence to components of cognitive-behavioral therapy for youth anxiety after training and consultation. *Prof Psychol Res Pr* 2017;48(1):54-61 [FREE Full text] [doi: [10.1037/pro0000100](https://doi.org/10.1037/pro0000100)] [Medline: [28603339](https://pubmed.ncbi.nlm.nih.gov/28603339/)]
56. Khademanian F, Aslani A, Bastani P. The effects of mobile apps on stress, anxiety, and depression: overview of systematic reviews. *Int J Technol Assess Health Care* 2020;37:e4. [doi: [10.1017/S0266462320002093](https://doi.org/10.1017/S0266462320002093)] [Medline: [33314997](https://pubmed.ncbi.nlm.nih.gov/33314997/)]
57. Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. *BMC Med* 2019;17(1):231 [FREE Full text] [doi: [10.1186/s12916-019-1461-z](https://doi.org/10.1186/s12916-019-1461-z)] [Medline: [31852455](https://pubmed.ncbi.nlm.nih.gov/31852455/)]
58. Dimidjian S, Barrera Jr M, Martell C, Muñoz RF, Lewinsohn PM. The origins and current status of behavioral activation treatments for depression. *Annu Rev Clin Psychol* 2011;7:1-38. [doi: [10.1146/annurev-clinpsy-032210-104535](https://doi.org/10.1146/annurev-clinpsy-032210-104535)] [Medline: [21275642](https://pubmed.ncbi.nlm.nih.gov/21275642/)]
59. Morriss RK, Faizal MA, Jones AP, Williamson PR, Bolton C, McCarthy JP. Interventions for helping people recognise early signs of recurrence in bipolar disorder. *Cochrane Database Syst Rev* 2007;2007(1):CD004854 [FREE Full text] [doi: [10.1002/14651858.CD004854.pub2](https://doi.org/10.1002/14651858.CD004854.pub2)] [Medline: [17253526](https://pubmed.ncbi.nlm.nih.gov/17253526/)]
60. East ML, Havard BC. Mental health mobile apps: from infusion to diffusion in the mental health social system. *JMIR Ment Health* 2015;2(1):e10 [FREE Full text] [doi: [10.2196/mental.3954](https://doi.org/10.2196/mental.3954)] [Medline: [26543907](https://pubmed.ncbi.nlm.nih.gov/26543907/)]
61. Meskó B, Drobni Z, Bényei É, Gergely B, Györfy Z. Digital health is a cultural transformation of traditional healthcare. *Mhealth* 2017;3:38 [FREE Full text] [doi: [10.21037/mhealth.2017.08.07](https://doi.org/10.21037/mhealth.2017.08.07)] [Medline: [29184890](https://pubmed.ncbi.nlm.nih.gov/29184890/)]

62. 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](https://doi.org/10.1371/journal.pone.0152285)] [Medline: [27073900](https://pubmed.ncbi.nlm.nih.gov/27073900/)]
63. Bentz D, Wang N, Ibach MK, Schicktanz NS, Zimmer A, Papassotiropoulos A, et al. Effectiveness of a stand-alone, smartphone-based virtual reality exposure app to reduce fear of heights in real-life: a randomized trial. *NPJ Digit Med* 2021;4(1):16 [FREE Full text] [doi: [10.1038/s41746-021-00387-7](https://doi.org/10.1038/s41746-021-00387-7)] [Medline: [33558625](https://pubmed.ncbi.nlm.nih.gov/33558625/)]
64. Pramana G, Parmanto B, Lomas J, Lindhiem O, Kendall PC, Silk J. Using mobile health gamification to facilitate cognitive behavioral therapy skills practice in child anxiety treatment: open clinical trial. *JMIR Serious Games* 2018;6(2):e9 [FREE Full text] [doi: [10.2196/games.8902](https://doi.org/10.2196/games.8902)] [Medline: [29748165](https://pubmed.ncbi.nlm.nih.gov/29748165/)]
65. D'Alfonso S. AI in mental health. *Curr Opin Psychol* 2020;36:112-117. [doi: [10.1016/j.copsyc.2020.04.005](https://doi.org/10.1016/j.copsyc.2020.04.005)] [Medline: [32604065](https://pubmed.ncbi.nlm.nih.gov/32604065/)]
66. Vaidyam AN, Wisniewski H, Halamka JD, Kashavan MS, Torous JB. Chatbots and conversational agents in mental health: a review of the psychiatric landscape. *Can J Psychiatry* 2019;64(7):456-464 [FREE Full text] [doi: [10.1177/0706743719828977](https://doi.org/10.1177/0706743719828977)] [Medline: [30897957](https://pubmed.ncbi.nlm.nih.gov/30897957/)]
67. Hariton E, Locascio JJ. Randomised controlled trials - the gold standard for effectiveness research: study design: randomised controlled trials. *BJOG* 2018;125(13):1716 [FREE Full text] [doi: [10.1111/1471-0528.15199](https://doi.org/10.1111/1471-0528.15199)] [Medline: [29916205](https://pubmed.ncbi.nlm.nih.gov/29916205/)]
68. Bauer M, Glenn T, Geddes J, Gitlin M, Grof P, Kessing LV, et al. Smartphones in mental health: a critical review of background issues, current status and future concerns. *Int J Bipolar Disord* 2020;8(1):2 [FREE Full text] [doi: [10.1186/s40345-019-0164-x](https://doi.org/10.1186/s40345-019-0164-x)] [Medline: [31919635](https://pubmed.ncbi.nlm.nih.gov/31919635/)]
69. Neff G, Nagy P. Agency in the digital age: using symbiotic agency to explain human-technology interaction. In: Papacharissi Z, editor. *A networked self and human augmentics, artificial intelligence, sentience*. New York, NY: Routledge; 2018:97-107.
70. Armstrong D. Actors, patients and agency: a recent history. *Sociol Health Illn* 2014;36(2):163-174. [doi: [10.1111/1467-9566.12100](https://doi.org/10.1111/1467-9566.12100)] [Medline: [24372176](https://pubmed.ncbi.nlm.nih.gov/24372176/)]
71. Kretzschmar K, Tyroll H, Pavarini G, Manzini A, Singh I, NeurOx Young People's Advisory Group. Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (Chatbots) in mental health support. *Biomed Inform Insights* 2019;11:1178222619829083 [FREE Full text] [doi: [10.1177/1178222619829083](https://doi.org/10.1177/1178222619829083)] [Medline: [30858710](https://pubmed.ncbi.nlm.nih.gov/30858710/)]
72. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: a conceptual overview and systematic review of relations with anxiety and depression psychopathology. *J Affect Disord* 2017;207:251-259. [doi: [10.1016/j.jad.2016.08.030](https://doi.org/10.1016/j.jad.2016.08.030)] [Medline: [27736736](https://pubmed.ncbi.nlm.nih.gov/27736736/)]
73. Denecke K, Hochreutener SL, Pöpel A, May R. Self-anamnesis with a conversational user interface: concept and usability study. *Methods Inf Med* 2018;57(5-06):243-252. [doi: [10.1055/s-0038-1675822](https://doi.org/10.1055/s-0038-1675822)] [Medline: [30875703](https://pubmed.ncbi.nlm.nih.gov/30875703/)]
74. Elias J. In 2016, users will trust health apps more than their doctors. *Forbes*. 2015. URL: <http://www.forbes.com/sites/jenniferelias/2015/12/31/in-2016-users-will-trust-health-apps-more-than-their-doctors/#4a244142d5f6> [accessed 2021-11-30]
75. Peters D, Calvo RA, Ryan RM. Designing for motivation, engagement and wellbeing in digital experience. *Front Psychol* 2018;9:797 [FREE Full text] [doi: [10.3389/fpsyg.2018.00797](https://doi.org/10.3389/fpsyg.2018.00797)] [Medline: [29892246](https://pubmed.ncbi.nlm.nih.gov/29892246/)]
76. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 2015;3(1):e27 [FREE Full text] [doi: [10.2196/mhealth.3422](https://doi.org/10.2196/mhealth.3422)] [Medline: [25760773](https://pubmed.ncbi.nlm.nih.gov/25760773/)]
77. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. *JMIR Ment Health* 2019;6(1):e11334 [FREE Full text] [doi: [10.2196/11334](https://doi.org/10.2196/11334)] [Medline: [30681968](https://pubmed.ncbi.nlm.nih.gov/30681968/)]
78. Abd-Alrazaq A, Safi Z, Alajlani M, Warren J, Househ M, Denecke K. Technical metrics used to evaluate health care chatbots: scoping review. *J Med Internet Res* 2020;22(6):e18301 [FREE Full text] [doi: [10.2196/18301](https://doi.org/10.2196/18301)] [Medline: [32442157](https://pubmed.ncbi.nlm.nih.gov/32442157/)]
79. Denecke K, Warren J. How to evaluate health applications with conversational user interface? *Stud Health Technol Inform* 2020;270:976-980. [doi: [10.3233/SHTI200307](https://doi.org/10.3233/SHTI200307)] [Medline: [32570527](https://pubmed.ncbi.nlm.nih.gov/32570527/)]
80. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. *J Med Internet Res* 2018;20(6):e10120 [FREE Full text] [doi: [10.2196/10120](https://doi.org/10.2196/10120)] [Medline: [29875087](https://pubmed.ncbi.nlm.nih.gov/29875087/)]
81. Lovell K, Bower P, Gellatly J, Byford S, Bee P, McMillan D, et al. Clinical effectiveness, cost-effectiveness and acceptability of low-intensity interventions in the management of obsessive-compulsive disorder: the Obsessive-Compulsive Treatment Efficacy randomised controlled Trial (OCTET). *Health Technol Assess* 2017;21(37):1-132 [FREE Full text] [doi: [10.3310/hta21370](https://doi.org/10.3310/hta21370)] [Medline: [28681717](https://pubmed.ncbi.nlm.nih.gov/28681717/)]

Abbreviations

CBT: cognitive behavioral therapy

mHealth: mobile health

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

RCT: randomized controlled trial

WHO: World Health Organization

Edited by A Mavragani; submitted 06.02.21; peer-reviewed by Y Khazaal, R Calvo, S Ng; comments to author 02.06.21; revised version received 27.07.21; accepted 28.12.21; published 10.03.22.

Please cite as:

Denecke K, Schmid N, Nüssli S

Implementation of Cognitive Behavioral Therapy in e-Mental Health Apps: Literature Review

J Med Internet Res 2022;24(3):e27791

URL: <https://www.jmir.org/2022/3/e27791>

doi: [10.2196/27791](https://doi.org/10.2196/27791)

PMID: [35266875](https://pubmed.ncbi.nlm.nih.gov/35266875/)

©Kerstin Denecke, Nicole Schmid, Stephan Nüssli. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 10.03.2022. 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.

Review

Improving the Development and Implementation of Audit and Feedback Systems to Support Health Care Workers in Limiting Antimicrobial Resistance in the Hospital: Scoping Review

Julia Keizer¹, PhD; Britt E Bente¹, MSc; Nashwan Al Naiemi^{2,3}, MD, PhD; Lisette JEW Van Gemert-Pijnen¹, PhD; Nienke Beerlage-De Jong⁴, PhD

¹Centre for eHealth and Wellbeing Research, Section of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

²Laboratorium Microbiologie Twente Achterhoek, Hengelo, Netherlands

³Department of Infection Prevention, Hospital Group Twente, Almelo/Hengelo, Netherlands

⁴Section of Health Technology and Services Research, Technical Medical Centre, University of Twente, Enschede, Netherlands

Corresponding Author:

Julia Keizer, PhD

Centre for eHealth and Wellbeing Research

Section of Psychology, Health and Technology

University of Twente

PO Box 217

Enschede, 7500 AE

Netherlands

Phone: 31 53 489 1289

Email: j.keizer@utwente.nl

Abstract

Background: For eHealth technologies in general and audit and feedback (AF) systems specifically, integrating interdisciplinary theoretical underpinnings is essential, as it increases the likelihood of achieving desired outcomes by ensuring a fit among eHealth technology, stakeholders, and their context. In addition, reporting on the development and implementation process of AF systems, including substantiations of choices, enables the identification of best practices and accumulation of knowledge across studies but is often not elaborated on in publications.

Objective: This scoping review aims to provide insights into the development and implementation strategies for AF systems for a real-world problem that threatens modern health care—antimicrobial resistance—and provide an interdisciplinary conceptual framework that can serve as a checklist and guidance for making informed choices in the development and implementation of future AF systems.

Methods: A scoping review was conducted by querying PubMed, Scopus, Web of Science, IEEE Xplore Digital Library, and Embase (≥2010) for studies describing either the development or implementation process, or both, of an AF system for antimicrobial resistance or infections in hospitals. Studies reporting only on effectiveness or impact were excluded. A total of 3 independent reviewers performed the study selection, and 2 reviewers constructed the conceptual framework through the axial and selective coding of often-used theories, models, and frameworks (TMFs) from the literature on AF and eHealth development and implementation. Subsequently, the conceptual framework was used for the systematic extraction and interpretation of the studies' descriptions of AF systems and their development and implementation.

Results: The search resulted in 2125 studies that were screened for eligibility, of which 12 (0.56%; 2012-2020) were included. These studies described the development and implementation processes heterogeneously in terms of study aims, study targets, target groups, methods, and theoretical underpinnings. Few studies have explicitly explained how choices for the development and implementation of AF systems were substantiated by the TMFs. The conceptual framework provided insights into what is reported on the development and implementation process and revealed underreported AF system constructs (eg, AF system design; engagement with the AF system; and comparison, goal setting, and action planning) and development and implementation (eg, champions) constructs.

Conclusions: This scoping review showed the current heterogeneous reporting of AF systems and their development and implementation processes and exemplified how interdisciplinary TMFs can (and should) be balanced in a conceptual framework to capture relevant AF systems and development and implementation constructs. Thereby, it provides a concrete checklist and

overall guidance that supports the professionalization and harmonization of AF system development and implementation. For the development and implementation of future AF systems and other eHealth technologies, researchers and health care workers should be supported in selecting and integrating TMFs into their development and implementation process and encouraged to explicitly report on theoretical underpinnings and the substantiation of choices.

(*J Med Internet Res* 2022;24(3):e33531) doi:[10.2196/33531](https://doi.org/10.2196/33531)

KEYWORDS

scoping review; audit and feedback; eHealth; development; implementation; antimicrobial resistance; antibiotic stewardship; infection control

Introduction

Background

Audit and feedback (AF) is a common reflective approach for various health care targets; however, the reported effects are small to moderate [1]. With the increase in electronically available data in health care, there is great potential for electronic AF systems [2]. The effectiveness of AF systems depends on the targeted behavior and the content, delivery, and context of AF and the system [1,3-6]. These constructs are often studied after AF system development and implementation to evaluate strategies and their ingredients for success [7-10]. In the literature, less attention has been paid to the development and implementation processes of AF systems [3], as is also common in the broader field of eHealth [11,12]. The development process of eHealth can refer to the entire iterative process of developing an eHealth technology, from predesign and design to implementation and (summative) evaluation [13]. However, in this study, we focus on the process from predesign and design (referred to hereafter as development) to implementation, including formative evaluation cycles. This allows us to focus on the early stages of implementation and development that are truly intertwined, as potential implementation issues (eg, limited eHealth skills) should be accounted for early in the development process to avoid well-known pitfalls of stakeholder and context disregard [14]. These phases are entwined by iterative formative evaluation cycles that provide ongoing information on how to improve both the eHealth technology and the development process taking [13].

Development and implementation are essential to gain a profound understanding of relevant stakeholders, their thinking and work processes, and their context (including implementation factors). Without this understanding, a misfit among technology, context, and people is likely to occur, which decreases the effectivity and efficiency of eHealth in practice [13]. It is crucial to consider these constructs from the start of the development and implementation process to avoid common pitfalls in current AF, such as top-down expert-driven audits with feedback at the hospital level rather than personalized, actionable feedback that supports health care workers (HCWs) in improving the quality and safety of health care [15,16].

The application of theories, models, and frameworks (TMFs) is advocated as an integral part of eHealth development and implementation as it identifies what works for whom, why, how, and when, likely resulting in eHealth technology that achieves the desired outcomes [17]. Colquhoun et al [18] and Tuti et al

[2] reported that only 9% (n=140) and 29% (n=7) of the included studies in their systematic reviews explicitly used theory to inform AF development and design. Therefore, implicit assumptions about AF working mechanisms and effective AF systems have driven AF development. Although these assumptions might hold true, they were not informed by theory [18,19], whereas there is a clear link between TMFs and eHealth intervention effectiveness [20,21].

To study the development and implementation of AF, this scoping review focuses on a real-world, wicked problem—antimicrobial resistance (AMR). AMR poses an increasing threat to human health and the durability of modern health care [22]. By 2050, AMR is expected to cause more yearly deaths worldwide than cancer currently does [23]. Antimicrobial and diagnostic stewardship programs and infection control programs form an integrated approach of AMR prevention measures (APMs) that aim to reduce and prevent the burden of AMR in hospitals [24]. Previous studies on HCWs' needs for APM support showed that changing HCWs' beliefs about their contribution to limiting AMR should be an important aim of APM strategies rather than merely focusing on raising AMR awareness or influencing ad hoc decisions [25,26]. To do so, learning through reflective cycles provides the opportunity to change HCWs' behaviors by giving them insight into their own behavior and improvement possibilities [15,27]. Therefore, AF for APM (APM-AF) is a promising strategy in the fight against AMR, although it is currently underused and understudied in the field of AMR [7].

There is a clear link between the use of TMFs and APM effectiveness [28-31], and because of the interdisciplinary nature of APM and eHealth, approaches for development and implementation are grounded in a plethora of TMFs [32]. In particular, APM-AF combines behavior change techniques [28-31], participatory eHealth development [33], human-centered and persuasive design [34-37], and improvement [38] and implementation [39] science. Moreover, TMFs have emerged for AF itself (eg, actionable feedback and feedback intervention theory [3-6]) and in the field of AMR (eg, integrated stewardship model [16,24,40]). Combining these TMFs into a conceptual framework that guides the development and implementation of APM-AF is challenging, and there is little guidance on how to create such interdisciplinary conceptual frameworks [41,42].

Objectives

There seems to be no standardized way of (theoretically) substantiating choices for and reporting on the development

and implementation of AF systems, which hinders the identification of best practices and knowledge accumulation [10,43]. Whereas other reviews on AF have mainly focused on the effectiveness of AF systems [1,2], this scoping review focuses on the development and implementation process to further harmonize and professionalize AF system development and implementation. The aim of this study is to gain insight into the development and implementation strategies for APM-AF systems by answering the following research questions:

1. What studies have been conducted so far to study the development and implementation of APM-AF systems?
2. What TMFs are used and described in studies on the development and implementation processes of APM-AF systems?
3. What information has been reported on APM-AF systems, and how are choices substantiated?
4. What information has been reported on the development and implementation processes of APM-AF systems, and how are choices substantiated?
5. What are the lessons learned for the development and implementation of APM-AF systems?

To allow for an evidence synthesis of information on the development and implementation of APM-AF, and because of the explorative aim and research questions in this study, a scoping review is preferred over a systematic literature review [44,45]. This scoping review provides an interdisciplinary conceptual framework that supports researchers, HCWs, and policy makers to make informed choices in APM-AF system development and implementation, with the aim of reducing the burden of AMR and improving the quality and safety of health care.

Methods

The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist was used to report on this scoping review without a prior registered review protocol [46]. This scoping review was designed by a multidisciplinary research team comprising AMR and eHealth experts.

Eligibility Criteria

Studies were included if (1) they described the development and implementation process of an AF system (including monitoring and surveillance systems), (2) the system provided feedback to HCWs, and (3) the system targeted AMR and infections in hospitals. A more elaborate description of development and implementation is provided in [Multimedia Appendix 1](#) [13]. We define AF systems as any system that comprises AF, wherein at least one of them (audit or feedback) is delivered or enhanced through the internet and related technologies [47]. As reporting on eHealth development and implementation processes is highly heterogeneous, there were no requirements for specific TMFs, methods, or eHealth technologies. Reviews and poster abstracts were excluded, as were studies outside the hospital setting. Evaluation studies that only reported on APM-AF effectiveness and impact without reporting on development and implementation were excluded.

However, constructs of formative evaluation (defined as “activities throughout the entire development process that provide ongoing information on how to improve the development process, outcomes of activities and eHealth technology” [13]) were included, as it is intertwined throughout the eHealth development and implementation process. A full list of eligibility criteria can be found in [Multimedia Appendix 1](#).

Information Sources, Search, and Selection of Evidence

A comprehensive and systematic literature search in PubMed, Scopus, Web of Science, IEEE Xplore Digital Library, and Embase was conducted without language restrictions. Only studies published in and after 2010 were considered, as both eHealth development and implementation and AMR and APM are rapidly advancing fields. Databases were queried by JK on November 2, 2020, except for Embase, which was queried on January 28, 2021. Together with an information specialist, AMR experts, and eHealth researchers, a structured query was constructed comprising the following terms: *audit* OR *monitor* OR *surveillance* AND *feedback* AND *develop** OR *implement** AND *infection* OR *antib** OR *antimicrobial* OR *resistance*. The results were uploaded to the Covidence web-based software platform (Veritas Health Innovation Ltd), where duplicates were removed. Sources of evidence were selected in a thorough screening process, including title and abstract screening and full-text screening by three researchers independently (JK, BB, and NBJ). After each round, conflicts were discussed until a consensus was reached.

Data Charting Process

To chart the data, JK created a data extraction form ([Multimedia Appendix 2](#) [2-5,18,43,48-50]) in Microsoft Excel. The general study characteristics extracted were first author, year, journal, country, study aims, targets and target groups, study design and methods, and theoretical underpinning. Given the heterogeneous study approaches and theoretical underpinnings of the included studies, a comprehensive overarching conceptual framework was needed to systematically analyze relevant constructs. The conceptual framework was grounded in often-used TMFs and best practices from various scientific perspectives on AF [3-5,18] and for the description, development, and implementation of eHealth interventions in general [2,43,48-50]. These TMFs and best practices were merged via an iterative axial and selective coding process by JK and NBJ. Thereby, overlapping and complementary constructs from various scientific perspectives were revealed. To structure all constructs, a distinction was made between constructs for APM-AF systems (n=41; research question 3) and constructs for development and implementation (n=35; research question 4).

The data extraction form was discussed within the research team, piloted, and iteratively refined throughout the assessment process. Note that this conceptual framework should be merely seen as an analytic framework for an organized way of thinking about and reporting on APM-AF systems from various perspectives and not as a theory explaining or predicting possible interrelations and outcomes.

Synthesis of Results

The main researcher (JK) read all full texts and systematically extracted the data using the data extraction form. Studies were scored with a + for a comprehensive, ~ for an incomplete, and - for a missing description for each data field. Descriptions were copied from the studies and further summarized per data field by describing variations among studies (ie, axial coding). In this process, data fields described by none of the studies were omitted ([Multimedia Appendix 2](#)), and other overlapping fields were combined. This reduced the number of data fields for APM-AF systems to 29. The translation and summarization of the extracted data into results were discussed in various rounds within the research team. Owing to the heterogeneity and qualitative nature of the included study designs, the richness

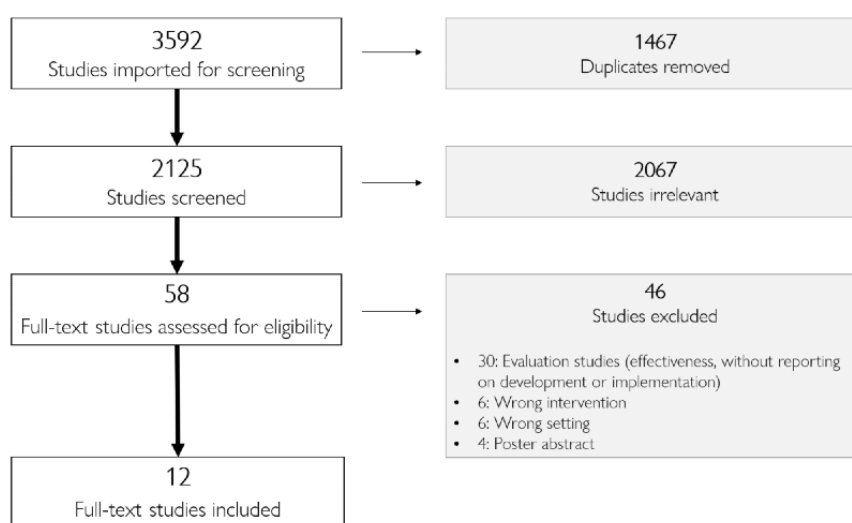
and relevance of the (contextual) information were believed to be more important than study quality. Therefore, no quality appraisal was performed [51].

Results

Study Selection

The literature search resulted in 3592 potentially relevant abstracts. Of the 3592 papers, after removing 1467 (40.84%) duplicates, 2125 (59.16%) unique titles and abstracts were assessed ([Figure 1](#)), which resulted in the eligibility assessment of 58 (1.61%) full texts. The main reasons for exclusion were a lack of information on development or implementation and evaluation studies (without reporting on development or implementation).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of included and excluded studies, including reasons for exclusion.



Current State of the Literature Addressing APM-AF Development and Implementation (Research Question 1)

Study Characteristics

In total, of the 58 papers, 12 (21%) were included in this review (2012-2020), mostly from PubMed, Scopus, and Web of

Science. Publications came from Northern American (6/12, 50%) or European (4/12, 33%) countries and Australia (2/12, 17%). Included studies stemmed from journals in various research fields (eg, infections or implementation science). Studies described APM-AF systems that were either in (preparation of) development or already implemented in practice, resulting in a wide variety of study aims, study targets, target groups, study designs, and used methods ([Table 1](#)).

Table 1. Study characteristics.

Author and country	Journal	Study aims	Study targets	Target group	Study design and methods	Theoretical underpinning
Boscart et al, Canada [52]	Implementation science	To identify nurses' and administrators' perceived barriers and facilitators to current HH ^a practices and the implementation of a new electronic monitoring technology for HH	ICP ^b and HAI ^c : HH (improving HH compliance)	Nurses and administrators	Qualitative: <ul style="list-style-type: none"> Semistructured key informant interviews 	Theoretical Domains Framework
Conway et al, United States [53]	The Joint Commission Journal on Quality and Patient Safety	To describe the implementation of an automated group monitoring and feedback system for promoting HH among HCWs ^d and report its impact on the frequency of HH at a community hospital	ICP and HAI: HH (increase HH frequency)	HCWs (eg, nurses and respiratory therapists), administrators, and managers	Multiple methods: <ul style="list-style-type: none"> Quantitative: before-and-after study on HH events per patient hour (outcome) Qualitative: focus groups 	Model of Actionable Feedback
Edmisten et al, United States [54]	American Journal of Infection Control	To describe the implementation of an electronic HH monitoring system in 3 community hospitals, including the challenges and drivers of success and the maintenance activities needed for continued improvements in compliance with HH practices	ICP and HAI: HH (improving HH compliance)	HCWs, staff, unit/department directors and, facility management	Multiple methods: <ul style="list-style-type: none"> Quantitative: after study (outcome measures on HH compliance after implementation) Qualitative: direct input from users/department and facility leaders, direct observation, and analysis of system-generated data and sharing of best practices between facilities 	None reported
Hysong et al, United States [55]	BMJ Quality and Safety	To describe how feedback intervention theory can be systematically applied in health care settings to design better feedback interventions	DSP ^e and HAI: to improve internal-medicine resident's and long-term care personnel's capacity to distinguish between asymptomatic bacteriuria and catheter-associated urinary tract infection	HCWs (eg, nurse practitioners and staff physicians)	Multiple methods: <ul style="list-style-type: none"> Quantitative: the Smither et al [56] 11-item scale for recipients' reactions to feedback Quantitative: chart monitoring (adherence to the treatment algorithm, specifically, rates of urine culture) of orders and inappropriate use of antibiotics 	Feedback Intervention Theory
James, Australia [57]	The Journal of Antimicrobial Chemotherapy	To design an audit tool that was appropriate for use in all Australian hospitals, suited to local user requirements, and included an assessment of the overall appropriateness of the prescription	ASP ^f : to improve the quality of patient care by reducing inappropriate and unnecessary use of antimicrobials (national level focus)	HCWs (eg, pharmacists and nurses)	Multiple methods: <ul style="list-style-type: none"> Quantitative: interrater reliability and validity tests and web-based questionnaire Qualitative: teleconference and direct input from users 	None reported
Jeanes et al, United Kingdom [58]	American Journal of Infection Control	To develop and implement an infection control performance and quality improvement data collection tool to meet the needs of large, acute health care providers and improve the credibility and use of infection control performance monitoring	ICP and HAI: to improve the credibility and use of infection control performance monitoring (beyond HH)	Not clearly described; ("auditors" and managers)	Multiple methods: <ul style="list-style-type: none"> Quantitative: questionnaires and intermittent validation Qualitative: day to day contacts with auditors, feedback from users via the IC-CQI^g data input system, discussion groups, and IC-CQI training sessions 	Pronovost Knowledge Translation Cycle and Barriers and Mitigation tool, double loop learning cycle, and Hexagon tool framework

Author and country	Journal	Study aims	Study targets	Target group	Study design and methods	Theoretical underpinning
Keizer, the Netherlands [59]	Lecture Notes in Computer Science	To describe how a bottom-up participatory development approach can improve the persuasive design of data-driven technologies for their end user (ie, HCWs) and within their context and describe how bottom-up participatory development is a necessary precondition for the development of persuasive data-driven technologies that foster sustainable implementation	DSP, ASP, and ICP: to optimize HCWs' diagnostic, antibiotic prescription and infection control behavior to limit AMR ^b	HCWs (eg, urologists and residents)	Multiple methods: <ul style="list-style-type: none"> Quantitative: questionnaire Qualitative: 2 focus groups (last focus group prototype based) 	CeHRes ¹ road map
Marques, Portugal [60]	BMC Medical Informatics and Decision Making	To develop a gamification solution that can provide HCWs real-time feedback on personal HH compliance	ICP and HAI: to create awareness regarding HCWs' HH compliance while trying to change their behaviors and optimize their performance	Nurses	Multiple methods (2 work iterations): <ul style="list-style-type: none"> Qualitative: preliminary experiments, simulations, field studies, and focus groups 	Design Science Research Methodology and gamification
Pakyz, United States [61]	American Journal of Infection Control	To identify the factors related to the implementation of ASP strategies	ASP: to optimize the use of antimicrobial agents, decrease AMR, and decrease rates of <i>Clostridium difficile</i> infection	ASP pharmacists and physicians	Multiple methods: <ul style="list-style-type: none"> Quantitative: survey Qualitative: semistructured telephone interviews 	None reported
Parker, Australia [62]	Journal of Clinical Nursing	To provide insights into the experiences of clinicians in implementing a multifaceted bundled urinary catheter care intervention (of which AP ^d is a considerable component)	HAI: the study aimed to reduce catheter use and duration of catheterization	Clinicians (eg, nurses and resident medical officers)	Qualitative: <ul style="list-style-type: none"> Postimplementation focus groups 	Intervention Description and Replication framework
Patel, United States [63]	Interdisciplinary Perspectives on Infectious Diseases	To describe the development and implementation of their AF intervention using a theoretical framework	ASP: to promote the judicious use of antibiotics	HCWs (eg, neonatologists and pediatric residents)	Multiple methods: <ul style="list-style-type: none"> Quantitative: retrospective observational study of antibiotic use and clinical vignette study Qualitative: ethnographic workflow study and 2 focus groups 	Model of Actionable Feedback
Power, United Kingdom [64]	International Journal for Quality in Health Care	To set up a low-cost pragmatic system to provide monthly data on 4 harms across care settings and produce measures that could be used locally for improvement but also aggregated to determine the burden of harm nationally	HAI: to reduce 4 high volume harms (safety outcomes), pressure ulcers, falls, urinary tract infection in patients with catheters, and venous thromboembolism	HCWs (eg, nurses and junior physicians)	Multiple methods: <ul style="list-style-type: none"> Quantitative: questionnaire survey (professional satisfaction) Qualitative: paper-based prototyping, formative evaluation by interaction with testers, web forum (including mail queries), regional leads, face-to-face meetings, and regional measurement workshops 	ProjectPplan Framework and Plan, Do, Study, Act Method

^aHH: hand hygiene.^bICP: infection control program.^cHAI: hospital-acquired infection.^dHCW: health care worker.^eDSP: diagnostic stewardship program.^fASP: antimicrobial stewardship program.

[§]IC-CQI: Infection Control Continuous Quality Improvement.

^hAMR: antimicrobial resistance.

ⁱCeHRes: Center for eHealth Research.

^jAF: audit and feedback.

Study Aims

Of the 12 studies, 4 (33%) primarily focused on development, 4 (33%) on implementation, and 4 (33%) described both development and implementation. However, development and implementation appeared to be undefined concepts, with *implementation* studies describing the development and design constructs and *development* studies paying attention to implementation constructs. Studies aimed at describing APM-AF system development focusing on (1) the integration of TMFs (eg, Feedback Intervention Theory), (2) AF content and presentation (eg, feedback gamification), or (3) technical aspects (eg, suitable badges for hand hygiene [HH] monitoring). In addition, studies focused on implementation barriers and facilitators.

Study Targets and Target Groups

Of the 12 studies, 11 (92%) focused on one of the APM (ie, diagnostic stewardship programs, antimicrobial stewardship programs, or infection control programs), whereas 1 (8%) targeted multiple APM. The target groups comprised a variety of HCWs (both frontline staff and AMR experts; 8/12, 67%) and in some studies, administrators and managers (4/12, 33%) as well.

Study Design and Methods

Most studies (10/12, 83%) used multiple methods, complementing quantitative (eg, questionnaires) with qualitative data (eg, observations, interviews, and focus groups). Approximately 17% (2/12) of the studies were fully qualitative, relying on interviews and focus groups.

TMFs for APM-AF Development and Implementation (Research Question 2)

Theoretical Underpinning Described by Studies

Most studies (9/12, 75%) described the theoretical underpinnings of their APM-AF system or study approach (Table 1). Approximately 17% (2/12) of the studies explicitly mentioned the use of Feedback Intervention Theory and the Model of Actionable Feedback to guide choices in the development and implementation of their study aims [55,63], whereas others mentioned TMFs in their Introduction or Methods section. AF TMFs (3/12, 25%; eg, Model of Actionable Feedback [53,55,63]) were used, as were TMFs, for developing, implementing, and evaluating interventions or technologies (5/12, 42%; eg, Center for eHealth Research road map [58-60,62,64]) and for identifying implementation barriers/facilitators (1/12, 8%; eg, Theoretical Domains Framework) [52]. Substantiations of choices on APM-AF systems were scarce; few studies substantiated their choices, which were either theory informed (eg, providing group-level feedback) or based on findings from the studies themselves (ie, revisions based on formative evaluation).

Conceptual Framework for APM-AF Development and Implementation

The conceptual framework, which is based on often-used TMFs and best practices for AF and eHealth interventions, is shown in Table 2 (APM-AF system constructs) and Table 3 (development and implementation constructs) and in Multimedia Appendix 2 in more detail.

Table 2. Conceptual framework: APM-AF^a system constructs (N=12)^b.

Constructs and subconstructs	Audit and feedback ^c				eHealth and interventions ^d			Implementation ^e		Studies, n (%)
	[18]	[3]	[5]	[4]	[48]	[43]	[2]	[49]	[50]	
Audit										
Auditees ^f					✓			✓		10 (83)
Main <i>input</i> ^b						✓				9 (75)
Feedback										
Feedback recipients ^f					✓		✓	✓		8 (67)
Main <i>output</i> ^f	✓					✓	✓			8 (67)
Level of individualization and specificity										
Feedback provided to individual, groups, or both ^f	✓		✓		✓	✓				11 (92)
Feedback is about the individual or team's own behaviors ^b	✓	✓	✓	✓						10 (83)
Feedback level specificity ^f	✓		✓	✓		✓				8 (67)
Comparison, goal setting, and action planning										
Comparison ^f	✓	✓		✓		✓	✓			8 (67)
Goal setting ^g	✓	✓		✓			✓	✓		5 (42)
Action planning ^g	✓	✓		✓		✓	✓	✓		4 (33)
Feedback framing and incentives										
Punitiveness ^b			✓	✓						6 (50)
Attack on self-identity ^f				✓					✓	4 (33)
Intrinsic and extrinsic reinforcement or incentives ^f				✓				✓		4 (33)
Timing										
Delivery timing ^f	✓	✓		✓		✓				8 (67)
Timeliness (frequency and continuous cycle) ^f	✓	✓	✓	✓	✓	✓	✓			11 (92)
Reminders ^f				✓		✓				3 (25)
APM-AF system										
Technology and materials										
Key features of the technology ^f						✓	✓		✓	11 (92)
Access ^b							✓			12 (100)
Materials ^b					✓	✓	✓			8 (67)
Human-system interaction										
Modes of feedback delivery ^f	✓	✓		✓	✓	✓	✓			9 (75)
Level of human involvement ^f							✓		✓	9 (75)
Engagement ^b				✓						6 (50)
Design										
Visual presentation strategies and cognitive load ^g	✓	✓	✓				✓	✓		9 (75)
User-guided experience ^g	✓	✓	✓	✓	✓	✓	✓	✓		4 (33)

Constructs and subconstructs	Audit and feedback ^c				eHealth and interventions ^d			Implementation ^e		Studies, n (%)
	[18]	[3]	[5]	[4]	[48]	[43]	[2]	[49]	[50]	
Data validity, d trust and credibility										
Data validity ^b		✓		✓						9 (75)
Trust and credibility ^f		✓		✓				✓		11 (92)
APM-AF as learning strategy										
Learning opportunities										
Reflective learning ^f	✓			✓				✓		5 (42)
Learning climate ^f				✓				✓		7 (58)
Additional strategies or procedures ^b					✓	✓	✓			12 (100)

^aAPM-AF: audit and feedback for antimicrobial resistance prevention measures.

^bUnique constructs (ie, where the various perspectives complement each other).

^cApproximately 72% of constructs theoretically underpinned by literature on audit and feedback.

^dApproximately 66% of constructs theoretically underpinned by literature on eHealth and interventions.

^eApproximately 41% of constructs theoretically underpinned by literature on implementation.

^fOverlapping constructs (constructs represented in 2 perspectives).

^gOverlapping constructs (constructs represented in all perspectives).

Table 3. Conceptual framework: APM-AF^a development and implementation constructs (N=12)^b.

Constructs and subconstructs	Audit and feedback ^c		eHealth and interventions ^d			Implementation ^e		Studies, n (%)		
	[18]	[3]	[5]	[4]	[48]	[43]	[2]		[49]	[50]
Stakeholders and roles										
Developer or research team ^b					✓					5 (42)
Pilot testers and involvement in development and implementation process ^f				✓	✓			✓		11 (92)
Leadership engagement ^b								✓	✓	6 (50)
Opinion leaders ^b								✓		3 (25)
Formally appointed internal implementation leaders ^b								✓		2 (17)
Champions ^b								✓		3 (25)
Target behavior and added value										
Target behavior, problem, and intervention										
Nature of the problem ^b									✓	12 (100)
Description of underlying behavior and decision processes ^b	✓			✓						12 (100)
Relevant sociocultural factors and comorbidities ^g									✓	8 (67)
Perceived need for behavior change ^g	✓							✓		4 (33)
Targeted behavior is likely to be amenable to feedback ^b	✓	✓		✓						6 (50)
Self-efficacy ^g		✓		✓				✓		3 (25)
Justify need for behavior change ^g	✓			✓					✓	10 (83)
Rationale and added value										
Rationale for using APM-AF ^g	✓				✓	✓				12 (100)
Desirability, efficacy, safety, and cost effectiveness ^g				✓				✓	✓	10 (83)
Relative advantage ^b								✓		10 (83)
Embedding in practice										
Implementation complexity and compatibility with target behavior and work processes										
Complexity of implementation process ^b								✓	✓	8 (67)
Technology supply model ^b									✓	8 (67)
Compatibility ^g				✓				✓	✓	11 (92)
Remove barriers ^g				✓		✓			✓	11 (92)
Opportunity costs (including additional efforts to use technology) ^g				✓				✓	✓	3 (25)
Available resources ^b								✓	✓	6 (50)
Inner and outer setting										
Structural characteristics ^b								✓	✓	1 (8)
Networks and communications ^b								✓	✓	2 (17)
Culture ^b								✓		3 (25)

Constructs and subconstructs	Audit and feedback ^c				eHealth and interventions ^d			Implementation ^e		Studies, n (%)
	[18]	[3]	[5]	[4]	[48]	[43]	[2]	[49]	[50]	
Patient needs and resources ^b								✓		1 (8)
Implementation										
Planning ^b								✓	✓	6 (50)
Executing ^b								✓	✓	5 (42)
Formative evaluation										
Intended use ^b					✓	✓				1 (8)
Actual use ^b					✓					3 (25)
Development process and formative evaluations ^g						✓		✓		12 (100)
Harms or unintended effects ^b						✓				4 (33)
Trialability ^b								✓		9 (75)
Revisions and updating ^g					✓			✓	✓	6 (50)
Replicability and digital preservation ^b								✓		1 (8)

^aAPM-AF: audit and feedback for antimicrobial resistance prevention measures.

^bUnique constructs (ie, where the various perspectives complement each other).

^cApproximately 32% of constructs theoretically underpinned by literature on audit and feedback.

^dApproximately 24% of constructs theoretically underpinned by literature on eHealth and interventions.

^eApproximately 74% of constructs theoretically underpinned by literature on implementation.

^fOverlapping constructs (constructs represented in all perspectives).

^gOverlapping constructs (constructs represented in 2 perspectives).

The construction of the comprehensive overarching conceptual framework revealed the added value of including multiple perspectives, as 48% of constructs were complementary (ie, covered by one of the 3 perspectives). Overlaps in the coverage of constructs between AF, eHealth, and implementation indicate the integrative nature of the development and implementation of the APM-AF system. Overlapping constructs occurred more often for APM-AF systems (4/29, 14%) than for APM-AF development and implementation (2/35, 6%). In the former, most constructs (21/29, 72% and 20/29, 69%, respectively) came from AF and eHealth literature, whereas in the latter, most constructs (28/35, 80%) came from the implementation

literature. Constructs underpinned by all 3 perspectives were not necessarily described by more studies (eg, goal setting; 5/12, 42%).

APM-AF System Constructs (Research Question 3)

Overview

Table 4 shows APM-AF system constructs (more details in Multimedia Appendix 3 [52-55,57-64]). Constructs could be categorized into four main codes—(1) audit, (2) feedback, (3) APM-AF system, and (4) APM-AF as a learning strategy—and are elaborated upon below. Table 4 also shows to what degree and by which studies these constructs were described.

Table 4. Constructs of APM-AF^a systems (N=12).

Constructs and subconstructs	Described by studies, n (%)			References
	Described elaborately and often substantiated	Partially described or constructed without elaboration or substantiation	Not described or substantiated	
Audit				
Auditees	10 (83)	1 (8)	1 (8)	[52-55,57,59,60,62-64]
Main input	9 (75)	3 (25)	0 (0)	[52-55,57,60,62-64]
Feedback				
Feedback recipients	8 (67)	3 (25)	1 (8)	[53-55,57,59,60,62-64]
Main output	8 (67)	3 (25)	1 (8)	[53-55,57-60,62,63]
Level of individualization and specificity				
Feedback provided to individual, groups, or both	11 (92)	1 (8)	0 (0)	[52-55,57,59-64]
Feedback about the individual or team's own behaviors	10 (83)	2 (17)	0 (0)	[52-55,59-64]
Specificity	8 (67)	1 (8)	3 (25)	[55,59,61,62,64]
Comparison, goal setting and action planning				
Comparison	8 (67)	0 (0)	4 (33)	[52,53,55,57,59,60,63,64]
Goal setting	5 (42)	1 (8)	6 (50)	[52-54,59]
Action planning	4 (33)	3 (25)	5 (42)	[55,58,59,61,62]
Framing and incentives				
Punitiveness	6 (50)	0 (0)	6 (50)	[53,55,59-61,63]
Attack on self-identity and cognitive influences	4 (33)	0 (0)	8 (67)	[53,58,61,63]
Intrinsic and extrinsic reinforcement or incentives	4 (33)	0 (0)	8 (67)	[52,55,60,61]
Timing				
Delivery timing	8 (67)	0 (0)	4 (33)	[52,54,60,61,64]
Timeliness	11 (92)	1 (8)	0 (0)	[52-55,57-60,62,63]
Reminders	3 (25)	0 (0)	9 (75)	[52,53,59]
APM-AF system				
Technology and materials				
Materials	11 (92)	1 (8)	0 (0)	[52-55,57-60,62-64]
Key features of the technology	12 (100)	0 (0)	0 (0)	[52-55,57-64]
Access	8 (67)	0 (0)	4 (33)	[53-55,59,60,62-64]
Human-system interaction				
Modes of feedback delivery	9 (75)	2 (17)	1 (8)	[52,53,55,59-64]
Level of human involvement	9 (75)	3 (25)	0 (0)	[53,55,57-59,61-64]
Engagement	6 (50)	0 (0)	6 (50)	[55,59,60,62-64]
Design				
Presentation strategies and cognitive load	9 (75)	1 (8)	2 (17)	[52-55,59,60,62-64]
User-guided experience	4 (33)	3 (25)	5 (42)	[55,59,60,64]
Validity and credibility				
Data validity	9 (75)	1 (8)	2 (17)	[52-55,57,58,60,61,64]

Constructs and subconstructs	Described by studies, n (%)			References
	Described elaborately and often substantiated	Partially described or constructed without elaboration or substantiation	Not described or substantiated	
Trust and credibility	11 (92)	0 (0)	1 (8)	[52,53,55,57-64]
APM-AF as a learning strategy				
Learning opportunities				
Reflective learning	5 (42)	0 (0)	7 (58)	[55,59-61]
Learning climate	7 (58)	0 (0)	5 (42)	[55,58,59,61,62,64]
Additional strategies or procedures	12 (100)	0 (0)	0 (0)	[52-55,57-64]

^aAPM-AF: audit and feedback for antimicrobial resistance prevention measures.

Audit

Auditees

The ones audited, or auditees, were described by most studies (10/12, 83%) and comprised frontline HCWs [52-55,57,59,60,62-64].

Main Input

Approximately 42% (5/12) APM-AF systems relied on automatically collected input (eg, electronic HH monitoring system) [52-54,60,63], whereas 33% (4/12) systems relied on manual input (eg, audit survey tool) [55,57,62,64].

Feedback

Feedback Recipients

Feedback recipients were described by most studies (8/12, 67%) and comprised auditees (ie, frontline HCWs) [55,60,62] and managers or administrators [53,54,59,63,64].

Main Output

Approximately 67% (8/12) of the studies described APM-AF output in terms of specific content (eg, process, structure, and outcome indicators) [53,54,58,60,62] or provided a general description of output (eg, quality of antibiotic treatment) [55,59,63].

Level of Individualization and Specificity

The level of feedback individualization was described by most studies (11/12, 92%). Feedback was provided on individual [55,61] or group level [53,62-64] or on both (ie, option to choose) [52,54,57,59,60]. Approximately 8% (2/12) of the studies explicitly justified their choice to provide group-level feedback only as individual feedback could be perceived as too threatening [53,63]. Feedback was provided to the auditees only [55,60-62], to the auditees and managers or administrators [52,54,59,63,64], or to managers or administrators only [53]. Feedback specificity was described by most studies (8/12, 67%). Feedback was provided on individual patient cases (mostly diagnostic and antimicrobial stewardship studies) [55,61], on aggregated group level (mostly infection control studies) [53,54,60], or on both individual and aggregated levels [59,62,64].

Comparison, Goal Setting, and Action Planning

Approximately 67% (8/12) of the studies described data comparison, either in terms of trends over time or benchmarks between groups and with other hospitals [52,57,59,60,63]. Approximately 33% (4/12) of the studies briefly mentioned goal setting and action planning. Goals were either derived from literature [52], based on current data [53,54], or described in terms of an HCW's need to discuss goals [59] but were not substantiated. Approximately 42% (5/12) of the studies mentioned action planning after feedback, but mostly in general terms (eg, feedback as a tool from which participants could make an actionable plan) [55,61,62], or as a separate follow-up strategy besides the APM-AF system, requiring additional information and human involvement [58,59,61,62].

Feedback Framing and Incentives

Some studies mentioned feedback framing in terms of punitiveness (6/12, 50%) or an attack on self-identity (3/12, 25%); however, few specified whether and how these constructs were incorporated in AF system design [53,55,58-61,63]. Approximately 8% (2/12) of the studies incorporated these constructs in their decision to focus only on group-level feedback [53,63], whereas 17% (2/12) studies emphasized nonconfrontational and informal language [60,61]. Intrinsic and extrinsic reinforcement and incentives were addressed in general terms [52,55,61] and more specifically by 8% (1/12) of the studies (eg, competition, win state, and rewards), including how these were implemented in the system (eg, presenting the highest score with a distinct color) [60].

Timing

Approximately 42% (5/12) of APM-AF systems made use of feedback at the point of care. This was provided via visual and auditory signals (for HH monitoring) [52,54] or a real-time performance dashboard [54,60,64]. Retrospective feedback was provided in 10 systems, with variable frequencies (daily, monthly, half yearly, and yearly) [52-55,57-60,62,63], sometimes with the possibility of continuously accessing the performance dashboard when needed [53,59,60]. Reminders were mentioned in 25% (3/12) of the studies [52,53,59].

APM-AF System

Technology and Materials

All studies described their (envisioned future) technologies, which ranged from audit tools (eg, Microsoft Excel or SurveyMonkey [Momentive Inc]) [57,58,61,62,64] to electronic monitoring systems (for HH) [52-54,60], interactive Microsoft PowerPoint presentations [55,63], and prototypes [59]. Access to the APM-AF systems was realized in interactive dashboards or Microsoft PowerPoint presentations with the possibility of sending customized reports via email [54,59,60,64], whereas, in 33% (4/12) of the studies, feedback recipients relied on managers or the research team (via email or face-to-face) for access to the APM-AF system [53,55,62,63].

Human-System Interaction

Approximately 17% (2/12) of the studies solely provided feedback via the APM-AF system [60,64], whereas 58% (7/12) of the studies also provided face-to-face feedback [52,53,55,59,61-63]. In addition, studies described the need for additional human involvement, for example, for (educational) meetings with AMR experts [53,57,59,62,63], and support in data processing [55,58,59,61,64]. Half of the studies described how they would engage the user with the APM-AF system via interactive feedback presentations [55,64], gamification [60], or with additional strategies (eg, creating an AF task force) [59,62,63].

Design

Design details about included graphs were described by 33% (4/12) of the studies [53,55,60,64], whereas 42% (5/12) of the studies broadly described the APM-AF system design [52,54,59,62,63]. One of the studies used a theory-informed design to ensure that their design matched task and user characteristics [55]. Approximately 33% (4/12) of the studies described an interactive and customizable AF system, wherein personalization was used to customize feedback to match end users' needs [55,59,60,64]. However, neither was this further

specified (eg, which parts are customizable) nor did it focus on user-guided experience (ie, how usability is ensured).

Validity and Credibility

Data validity was addressed in terms of raised concerns by study participants [52-54,61], (planned) data validation activities [55,57,58,64], or technical constructs [60]. The trust in and credibility of the APM-AF system was addressed by describing study participants' perceptions [52,53,58-62] or (planned) activities to improve the trust in and credibility of the system [55,57,63,64].

APM-AF as a Learning Strategy

Learning Opportunities

Approximately 33% (4/12) of the studies described reflective learning (ie, personal reflections on one's behavior and APM performance) as a result and strength of APM-AF systems [55,59-61]. Furthermore, APM-AF systems (6/12, 50%) were described as a (potential) facilitator for interactive discussions and communication between HCWs and AMR experts, mostly in existing meetings [55,58,59,61,62,64].

Additional Strategies or Procedures

All studies described additional strategies, both for the study (eg, webinar to explain how to use the tool) and for the APM-AF system in practice (eg, creating an AF task force) [52-55,57-64].

APM-AF Development and Implementation Constructs (Research Question 4)

Overview

Table 5 shows the APM-AF development and implementation constructs (Multimedia Appendix 4 [52-55,57-64] provides more details). Constructs could be categorized into four main codes—(1) stakeholders and roles, (2) target behavior and added value, (3) embedding in practice and (4) formative evaluation—and are elaborated upon below. Table 5 also shows to what degree and by which studies these constructs were described.

Table 5. APM-AF^a development and implementation constructs.

Constructs and subconstructs	Described by studies, n (%)			References
	Described elaborately and often substantiated	Partially described or constructed without elaboration or substantiation	Not described or substantiated	
Stakeholders and roles				
Stakeholders				
Developer or research team	5 (42)	4 (33)	3 (25)	[53,55,57,59,61]
Pilot testers and involvement in development and implementation process	11 (92)	2 (17)	1 (8)	[52-55,57,59-64]
Leadership engagement	6 (50)	2 (17)	4 (33)	[54,58,61-63]
Opinion leaders	3 (25)	0 (0)	9 (75)	[58,61,62]
Formally appointed internal implementation leaders	2 (17)	0 (0)	10 (83)	[61,62]
Champions	3 (25)	1 (8)	8 (67)	[52,61,62]
Target behavior and added value				
Target behavior, problem, and intervention				
Description of underlying behavior and decision processes	8 (67)	2 (17)	2 (17)	[52,53,55,57-59,62,63]
Nature of the problem	12 (100)	0 (0)	0 (0)	[52-55,57-64]
Relevant sociocultural factors and comorbidities	12 (100)	0 (0)	0 (0)	[52-55,57-64]
Tension for behavior change	4 (33)	1 (8)	7 (58)	[52,53,55,60]
Targeted behavior is likely to be amenable to feedback	6 (50)	0 (0)	6 (50)	[52,53,55,60-62]
Self-efficacy	3 (25)	0 (0)	9 (75)	[52,53,57]
Justify need for behavior change	10 (83)	0 (0)	2 (17)	[52-55,58-62,64]
Rationale and added value				
Rationale for using AF ^b	12 (100)	0 (0)	0 (0)	[52-55,57-64]
Desirability, efficacy, safety, and cost-effectiveness	10 (83)	0 (0)	2 (17)	[52,53,55,57-62,64]
Relative advantage	10 (83)	0 (0)	2 (17)	[52,53,55,57-62,64]
Embedding in practice				
Implementation complexity and compatibility with target behavior and work processes				
Complexity of implementation process	8 (67)	1 (8)	3 (25)	[54,58-64]
Technology supply model	8 (67)	0 (0)	4 (33)	[53,54,58,60-64]
Compatibility	11 (92)	1 (8)	0 (0)	[52,54,55,57-64]
Remove barriers	11 (92)	0 (0)	1 (8)	[52,54,55,57-64]
Opportunity costs (including additional efforts to use technology)	3 (25)	1 (8)	8 (67)	[58,60,62]
Available resources	6 (50)	2 (17)	4 (33)	[52,57,58,61-63]
Inner and outer setting				
Structural characteristics	1 (8)	0 (0)	11 (92)	[62]
Networks and communications	2 (17)	0 (0)	10 (83)	[61,62]
Culture	3 (25)	3 (25)	6 (50)	[59,61,63]
Patient needs and resources	1 (8)	1 (8)	10 (83)	[54]
Implementation				

Constructs and subconstructs	Described by studies, n (%)			References
	Described elaborately and often substantiated	Partially described or constructed without elaboration or substantiation	Not described or substantiated	
Planning	6 (50)	0 (0)	6 (50)	[54,58-60,62,64]
Execution	5 (42)	0 (0)	7 (58)	[54,58,60,62,64]
Formative evaluation				
APM-AF system use				
Intended use	1 (8)	1 (8)	10 (83)	[64]
Actual use	3 (25)	2 (17)	7 (58)	[58,60,64]
Development process and formative evaluations	11 (92)	1 (8)	0 (0)	[52-55,57-60,62-64]
Harms or unintended effects	4 (33)	0 (0)	8 (67)	[54,61,63,64]
Trialability	9 (75)	1 (8)	2 (17)	[52-55,57,59,60,62,64]
Revisions and updating	6 (50)	1 (8)	5 (42)	[52-54,57,60,64]
Replicability and digital preservation	1 (8)	1 (8)	10 (83)	[64]

^aAPM-AF: audit and feedback for antimicrobial resistance prevention measures.

^bAF: audit and feedback.

Stakeholders and Roles

Research Team

Approximately 42% (5/12) of the studies described their research team, comprising multidisciplinary stakeholders (eg, HCWs, AMR experts, biostatisticians, and researchers) [53,55,57,59,61]. The research team compositions were substantiated (eg, having a multidisciplinary mix in the team [53,55,57,59,61] and experience with specific research methods [53,55,61]).

Pilot Testers and Involvement in Development and Implementation Process

Pilot testers were described by 92% (11/12) of the studies [52-55,57,59-64] and were predominantly selected for their occupational function [52,55,57,59,60,62-64], whereas other details about stakeholders' expertise and background were hardly described [52,53,57]. Stakeholder involvement was realized by including stakeholders (eg, HCWs and AMR experts) in the research team [62,63] and by involving them in pilot tests and formative evaluations [57,59,60,62-64]. Leadership engagement was mentioned by half of the studies as facilitator for successful implementation [54,58,61-63], whereas stakeholder engagement through champions or opinion leaders was mentioned less often (4/12, 33%) [52,58,61,62].

Target Behavior and Added Value

Target Behavior, Problem, and Intervention

The nature of the problem and relevant sociocultural factors were addressed by all studies [52-55,57-64]. Most studies (8/12, 67%) provided a description of underlying behavior and decision processes, either shortly in the article's introduction [53,57] or in a prior study [52,55,58,59,62,63]. Some studies paid attention to whether there was tension for behavior change [52,53,55,60], whether the targeted behavior is likely to be amenable to

feedback [52,53,55,60-62], and self-efficacy of feedback recipients' (ie, feeling capable and responsible for behavior improvement) [52,53,57]. The need for behavior change was justified by pointing out flaws in current behaviors [52-55,58-62,64].

Rationale and Added Value

All studies described the rationale and added value of the APM-AF [52-55,57-64]. Approximately 58% (7/12) of the studies described recommendations from health authorities (eg, World Health Organization) or AF as a widely used intervention in health care in general as reasons for developing and implementing an APM-AF system [53-55,57,59,61,63]. Other studies (5/12, 42%) explained how APM-AF could solve problems and inefficiencies in the current situation [52,58,60,62,64]. The added value was described both in terms of expectations (eg, improving the efficiency of audits) and experiences (eg, feedback motivated to change behavior) [52,53,55,57-62,64].

Embedding in Practice

Implementation Complexity and Compatibility With Target Behavior and Work Processes

Most studies (8/12, 67%) described constructs related to implementation complexity, including required organizational configurations and dependability on suppliers for customizations in terms of expected or experienced implementation barriers [53,54,58-64]. One of the studies specifically reflected on the duration and effort of the entire implementation process [58]. Approximately all studies (11/12, 92%) described constructs regarding the compatibility between the APM-AF system and stakeholders' needs and existing workflows and expected or experienced implementation facilitators and barriers [52,54,55,57-64]. Opportunity costs were described by a few studies (3/12, 25%) [58,60,62], including negative experiences with the required additional efforts to use the APM-AF

(including education and collecting data) [60,62]. A lack of resources was described in terms of staffing, time, and materials [57,58,61-63].

Inner and Outer Setting

Few studies (4/12, 33%) paid attention to the inner setting, expressing the need for a collaborative environment and open culture, in which the quality of their work can be discussed safely [59,61,63]. One of the studies described increased patient involvement as a result of visible APM-AF systems [54].

Implementation Planning and Execution

Approximately 50% (6/12) of the studies addressed implementation planning, of which 83% (5/6) also reflected on execution [54,58-60,62,64]. Studies mostly reported longer implementation processes than initially planned because of hospital workflow conflicts, personnel availability, and other confounding factors (including technical problems and resistance from stakeholders).

Formative Evaluation

APM-AF System Use

Intended and actual use of the APM-AF system was hardly (3/12, 25%) described, either as the maximum time HCWs should spend on filling out the audit tool [64] or as how often and complete the audit tool was filled in [58]. In addition, one of the studies used Google Analytics to measure users' interactions with gamification parts [60].

Development Process and Formative Evaluations

The development process and methods used for the formative evaluations were described in all studies [52-55,57-64] and elaborated in this paper's *Study Characteristics* section.

Harms or Unintended Effects

Approximately 33% (4/12) of the studies described whether and how harms and unintended effects were monitored during the development and implementation process (in general terms or with specific examples) [54,61,63,64].

Trialability and Revisions and Updating

Approximately 75% (9/12) of the studies described trialability and revisions and updates in terms of several testing rounds [52-55,57,59,60,62,64]. However, only 50% (6/12) of the studies subsequently described, either specifically [57,60,64] or broadly [52-54], how the findings from the testing rounds were incorporated in the design or implementation of the APM-AF system (eg, use of better beacons).

Replicability and Digital Preservation

One of the studies published their APM-AF system on the web with additional information (eg, web forum) [64].

Discussion

Principal Findings

This scoping review aimed to provide insights into strategies and theoretical underpinnings for AF system development and implementation from a sociotechnical perspective. Of the 2125 studies found in the search, 12 (0.56%) studies were included

describing the development and implementation of their AF systems heterogeneously in terms of study aims, AF targets, and development and implementation strategies. Approximately 17% (2/12) of the studies explicitly aimed to illustrate how TMFs could guide choices in AF system development and implementation. A comprehensive interdisciplinary conceptual framework, based on overlapping and complementary constructs from TMFs from AF, eHealth and interventions, and implementation literature, was created to compare the studies.

Lessons Learned for the Development and Implementation of APM-AF Systems (Research Question 5)

In this discussion, research question 5 is answered by providing lessons learned from reflecting upon our findings for theoretical underpinnings of and reporting on AF, AF systems, and their development and implementation.

Theoretical Underpinnings for AF

With health- and health care-related problems becoming increasingly complex, interdisciplinary theoretical integration to combine different perspectives is inevitable and pivotal to grasp the complexity of real-world problems [65]. This study showed the added value of considering and combining AF, eHealth, and implementation literature for studying AF systems. AF literature covered mostly AF system constructs (21/29, 72%), whereas relevant development and implementation constructs were hardly covered (12/35, 34%). Therefore, studies using only AF literature might miss important development and implementation constructs, such as stakeholder roles (eg, leadership engagement and champions) and embedding in practice (eg, implementation complexity and setting), that influence AF effectivity and efficiency [66,67]. Therefore, we argue that TMFs for AF and for development and implementation should be balanced, as exemplified in our interdisciplinary conceptual framework, and matched with studies' specific research objectives, methods, and context (eg, setting and participants) [68].

However, selecting and combining the best-fitting TMFs remains a challenge [69]. Well-known examples of combined frameworks exist in implementation science (eg, Theoretical Domains Framework [70]); however, little information is provided about how constructs were combined. Overall, there is little guidance on the selection and integration of interdisciplinary TMFs [71]. Evolving initiatives that create shared languages across disciplines and theories (eg, CohenMiller and Pate [72]) and provide criteria for theory selection (eg, Birken et al [73]) are encouraged.

Reporting on AF Systems

This scoping review resulted in an overview of constructs for APM-AF systems (Table 4), enriching the AF best practices from Colquhoun et al [18] with constructs of audit (eg, auditees and audit input), feedback framing and incentives, and AF system constructs (eg, technology and materials, human-system interaction, and data validity and trust and credibility). For replicability and using the framework in actual development and implementation, reporting about the audit input and what technology and materials were used is important. Furthermore,

data validity and credibility was deemed as one of the 5 most important aspects for AF in a recent study [42].

In our view, the 2 constructs that were underreported in the included studies require attention in future studies. First, we observed quite broad descriptions of AF system design, with a lack of attention to functional (ie, what can the AF system do) and visual (ie, how efficiently and effectively information is presented to users) design, and engagement with the AF system. However, these constructs are important for how an AF system is received and used in practice [74]. The lack of design aspects and considerations of engagement might reflect the neglect of eHealth system characteristics (such as design) and engagement as active influencers for eHealth effectivity [75]. Second, comparison, goal setting, and action planning were hardly described in the included studies. A lack of substantiations for comparisons was also reported by a review on clinical performance comparators for AF on various clinical topics [76], whereas the underuse of goal setting and action planning was also found by a systematic review on behavior change interventions for APM in hospitals [28]. These 3 constructs were derived from all included theoretical perspectives and are common behavior change techniques [77], suggesting that they require and deserve more attention in future studies.

Reporting on Development and Implementation of AF Systems

Of the full-text studies assessed for eligibility, most studies (30/58, 52%) were excluded as they focused primarily on effectiveness and did not sufficiently report on development and implementation. The inclusion of 12 studies in full-text might seem little; however, we believe this is exemplary of the tendency in (APM and eHealth) research to publish more about effectiveness than about the development and implementation process [3,11,12]. Therefore, in future AF system studies, but also for other eHealth technologies, considering development and implementation as influencers of the effectivity and efficiency of eHealth in practice has yet to gain ground [13]. During the construction of the conceptual framework, the interwovenness of development, implementation, and formative evaluation became apparent through the many overlaps observed. This resonates with best practices in eHealth development and implementation, which state that implementation is integrated with development and requires an iterative and holistic approach [78]. Therefore, next to reporting on evaluation, studies should report on both the constructs for AF systems and for development and implementation.

There is no single right development and implementation approach because of the many variations in APM and AF objectives, stakeholders, technologies, and settings [79]. Therefore, the constructed conceptual framework should be seen as a checklist that provides general guidance for potentially interesting constructs to consider; it remains at the discretion of researchers and developers which and how constructs are incorporated in their study. At a minimum, we propose to reflect upon relevant stakeholders and their roles, implementation complexity, compatibility with target behavior, and work processes, including the added value of AF and formative evaluation. Overall, supporting researchers and HCWs in

selecting and integrating TMFs into their development and implementation process and promoting the explicit reporting of the theoretical underpinning and substantiation of choices are highly encouraged [41].

The constructed conceptual framework can be used in future studies to ensure a comprehensive view of AF for the transparency and replicability of individual studies. Therefore, we recommend using the conceptual framework as a checklist and adding it (including substantiation of choices) as a supplementary material in future publications. Furthermore, findings from this study can be used to improve the professionalization and harmonization of AF studies, which is important considering the increasing use of AF principles in learning health care systems [80]. The lack of attention to factors that support the development of learning health care systems (eg, organizational culture and cooperation between HCWs and researchers) is recognized as an important barrier to the widespread adoption of learning health care systems [81]. As these aspects are included in the constructed conceptual framework, it might help future researchers and developers to explicitly consider and integrate these constructs in their AF or learning system.

Limitations and Strengths

This scoping review has several limitations. Although a comprehensive search query was used to search the most important databases for health research, we only included peer-reviewed published research and excluded evaluation studies. As a result, it might be possible that relevant findings were omitted (eg, from gray literature). Two systematic reviews on AF for various health targets were screened to ensure that no relevant publications were missed [1,2]. Another limitation is that evaluation studies were excluded from this review to highlight constructs relevant to the development and implementation, whereas evaluation is critical to know whether an intervention was successful. Therefore, it will be an important next step to evaluate AF systems in terms of processes (eg, improved HH), clinical outcomes (eg, reduced number of infections and decreased AMR), and technological outcomes (eg, APM-AF system use and persuasiveness) [82]. Finally, data extraction relied on the subjective interpretation of the constructs from the included publications by 1 researcher. However, the conceptual framework (Table 2 and Table 3) provided a thorough base for systematic and structured assessment, and the findings were iteratively discussed and revised throughout the review process.

Conclusions

This scoping review provides novel insights into the theoretical underpinning of and reporting on the development and implementation of AF systems while demonstrating how a comprehensive conceptual framework can be created and used and a valuable means for capturing relevant constructs from heterogeneous studies with varying theoretical underpinnings. Few studies have explicitly described how choices for AF systems and their development and implementation processes were substantiated by TMFs. The interdisciplinary conceptual framework developed in this study is a first step toward the professionalization and harmonization of AF development and

implementation with a sociotechnical approach. It provides guidance and a comprehensive checklist to guide researchers, HCWs, and policy makers in making informed choices in the development and implementation of AF systems, with the aim of further improving the quality and safety of health care.

Acknowledgments

This research was supported by the INTERREG-VA (grant 202085) funded project EurHealth-1Health, which is part of a Dutch–German cross-border network supported by the European Commission; the Dutch Ministry of Health, Welfare and Sport; the Ministry of Economy, Innovation, Digitalization and Energy of the German Federal State of North Rhine-Westphalia; and the Ministry for National and European Affairs and Regional Development of Lower Saxony.

Authors' Contributions

JK and NBJ conceptualized the manuscript. JK, BEB, and NBJ performed the data analysis. JK, NBJ, BEB, and LJEWCGP discussed the assessment form and results. NBJ, NAN, and LJEWCGP supervised the work. JK wrote the original draft, and all authors reviewed and edited the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Eligibility criteria.

[\[DOCX File, 14 KB - jmir_v24i3e33531_app1.docx\]](#)

Multimedia Appendix 2

Data extraction form.

[\[XLSX File \(Microsoft Excel File\), 56 KB - jmir_v24i3e33531_app2.xlsx\]](#)

Multimedia Appendix 3

Audit and feedback for antimicrobial resistance prevention measures systems.

[\[XLSX File \(Microsoft Excel File\), 252 KB - jmir_v24i3e33531_app3.xlsx\]](#)

Multimedia Appendix 4

Audit and feedback for antimicrobial resistance prevention measures development and implementation.

[\[XLSX File \(Microsoft Excel File\), 243 KB - jmir_v24i3e33531_app4.xlsx\]](#)

References

1. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev* 2012 Jun 13(6):CD000259. [doi: [10.1002/14651858.CD000259.pub3](https://doi.org/10.1002/14651858.CD000259.pub3)] [Medline: [22696318](https://pubmed.ncbi.nlm.nih.gov/22696318/)]
2. Tuti T, Nzinga J, Njoroge M, Brown B, Peek N, English M, et al. A systematic review of electronic audit and feedback: intervention effectiveness and use of behaviour change theory. *Implement Sci* 2017 May 12;12(1):61 [FREE Full text] [doi: [10.1186/s13012-017-0590-z](https://doi.org/10.1186/s13012-017-0590-z)] [Medline: [28494799](https://pubmed.ncbi.nlm.nih.gov/28494799/)]
3. Ivers NM, Sales A, Colquhoun H, Michie S, Foy R, Francis JJ, et al. No more 'business as usual' with audit and feedback interventions: towards an agenda for a reinvigorated intervention. *Implement Sci* 2014 Jan 17;9:14 [FREE Full text] [doi: [10.1186/1748-5908-9-14](https://doi.org/10.1186/1748-5908-9-14)] [Medline: [24438584](https://pubmed.ncbi.nlm.nih.gov/24438584/)]
4. Colquhoun HL, Carroll K, Eva KW, Grimshaw JM, Ivers N, Michie S, et al. Advancing the literature on designing audit and feedback interventions: identifying theory-informed hypotheses. *Implement Sci* 2017 Sep 29;12(1):117 [FREE Full text] [doi: [10.1186/s13012-017-0646-0](https://doi.org/10.1186/s13012-017-0646-0)] [Medline: [28962632](https://pubmed.ncbi.nlm.nih.gov/28962632/)]
5. Hysong SJ, Best RG, Pugh JA. Audit and feedback and clinical practice guideline adherence: making feedback actionable. *Implement Sci* 2006 Apr 28;1:9 [FREE Full text] [doi: [10.1186/1748-5908-1-9](https://doi.org/10.1186/1748-5908-1-9)] [Medline: [16722539](https://pubmed.ncbi.nlm.nih.gov/16722539/)]
6. Kluger AN, DeNisi A. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol Bulletin* 1996 Mar;119(2):254-284. [doi: [10.1037/0033-2909.119.2.254](https://doi.org/10.1037/0033-2909.119.2.254)]
7. Davey P, Marwick CA, Scott CL, Charani E, McNeil K, Brown E, et al. Interventions to improve antibiotic prescribing practices for hospital inpatients. *Cochrane Database Syst Rev* 2017 Feb 09;2:CD003543 [FREE Full text] [doi: [10.1002/14651858.CD003543.pub4](https://doi.org/10.1002/14651858.CD003543.pub4)] [Medline: [28178770](https://pubmed.ncbi.nlm.nih.gov/28178770/)]

8. Neo JR, Niederdeppe J, Vielemeyer O, Lau B, Demetres M, Sadatsafavi H. Evidence-based strategies in using persuasive interventions to optimize antimicrobial use in healthcare: a narrative review. *J Med Syst* 2020 Feb 10;44(3):64. [doi: [10.1007/s10916-020-1531-y](https://doi.org/10.1007/s10916-020-1531-y)] [Medline: [32040649](https://pubmed.ncbi.nlm.nih.gov/32040649/)]
9. Livorsi DJ, Heintz B, Jacob JT, Krein SL, Morgan DJ, Perencevich EN. Audit and feedback processes among antimicrobial stewardship programs: a survey of the society for healthcare epidemiology of America research network. *Infect Control Hosp Epidemiol* 2016 Jun 10;37(6):704-706. [doi: [10.1017/ice.2016.57](https://doi.org/10.1017/ice.2016.57)] [Medline: [26961763](https://pubmed.ncbi.nlm.nih.gov/26961763/)]
10. Baysari MT, Lehnbohm EC, Li L, Hargreaves A, Day RO, Westbrook JI. The effectiveness of information technology to improve antimicrobial prescribing in hospitals: a systematic review and meta-analysis. *Int J Med Inform* 2016 Aug;92:15-34. [doi: [10.1016/j.ijmedinf.2016.04.008](https://doi.org/10.1016/j.ijmedinf.2016.04.008)] [Medline: [27318068](https://pubmed.ncbi.nlm.nih.gov/27318068/)]
11. van Limburg M, van Gemert-Pijnen JEW, Nijland N, Ossebaard HC, Hendrix RM, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. *J Med Internet Res* 2011 Dec 28;13(4):e124 [FREE Full text] [doi: [10.2196/jmir.1674](https://doi.org/10.2196/jmir.1674)] [Medline: [22204896](https://pubmed.ncbi.nlm.nih.gov/22204896/)]
12. Kip H, Kelders SM, Bouman YH, van Gemert-Pijnen JEW. The importance of systematically reporting and reflecting on eHealth development: participatory development process of a virtual reality application for forensic mental health care. *J Med Internet Res* 2019 Aug 19;21(8):e12972 [FREE Full text] [doi: [10.2196/12972](https://doi.org/10.2196/12972)] [Medline: [31429415](https://pubmed.ncbi.nlm.nih.gov/31429415/)]
13. van Gemert-Pijnen JEW, Kelders SM, Kip H, Sanderman R. *eHealth Research, Theory and Development: A Multi-Disciplinary Approach*. London: Routledge; 2018.
14. van Gemert-Pijnen JEW, 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](https://doi.org/10.2196/jmir.1672)] [Medline: [22155738](https://pubmed.ncbi.nlm.nih.gov/22155738/)]
15. Keizer J, Beerlage-De Jong N, Al Naiemi N, van Gemert-Pijnen JEW. Finding the match between healthcare worker and expert for optimal audit and feedback on antimicrobial resistance prevention measures. *Antimicrob Resist Infect Control* 2020 Aug 05;9(1):125 [FREE Full text] [doi: [10.1186/s13756-020-00794-7](https://doi.org/10.1186/s13756-020-00794-7)] [Medline: [32758300](https://pubmed.ncbi.nlm.nih.gov/32758300/)]
16. Core elements of hospital antibiotic stewardship programs. Centers for Disease Control and Prevention. 2019. URL: <https://www.cdc.gov/antibiotic-use/core-elements/hospital.html> [accessed 2022-03-01]
17. 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](https://doi.org/10.1146/annurev.publhealth.012809.103604)] [Medline: [20070207](https://pubmed.ncbi.nlm.nih.gov/20070207/)]
18. Colquhoun H, Michie S, Sales A, Ivers N, Grimshaw JM, Carroll K, et al. Reporting and design elements of audit and feedback interventions: a secondary review. *BMJ Qual Saf* 2017 Jan 25;26(1):54-60. [doi: [10.1136/bmjqs-2015-005004](https://doi.org/10.1136/bmjqs-2015-005004)] [Medline: [26811541](https://pubmed.ncbi.nlm.nih.gov/26811541/)]
19. Gardner B, Whittington C, McAteer J, Eccles MP, Michie S. Using theory to synthesise evidence from behaviour change interventions: the example of audit and feedback. *Soc Sci Med* 2010 May;70(10):1618-1625. [doi: [10.1016/j.socscimed.2010.01.039](https://doi.org/10.1016/j.socscimed.2010.01.039)] [Medline: [20207464](https://pubmed.ncbi.nlm.nih.gov/20207464/)]
20. 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](https://doi.org/10.2196/jmir.1376)] [Medline: [20164043](https://pubmed.ncbi.nlm.nih.gov/20164043/)]
21. Morrison LG. Theory-based strategies for enhancing the impact and usage of digital health behaviour change interventions: a review. *Digit Health* 2015;1:2055207615595335 [FREE Full text] [doi: [10.1177/2055207615595335](https://doi.org/10.1177/2055207615595335)] [Medline: [29942544](https://pubmed.ncbi.nlm.nih.gov/29942544/)]
22. *Global Action Plan on Antimicrobial Resistance*. Geneva: World Health Organization; 2015.
23. O'Neill J. *Antimicrobial Resistance: Tackling A Crisis for the Health and Wealth of Nations*. London: Review on Antimicrobial Resistance; 2014.
24. Dik JH, Poelman R, Friedrich AW, Panday PN, Lo-Ten-Foe JR, van Assen S, et al. An integrated stewardship model: antimicrobial, infection prevention and diagnostic (AID). *Future Microbiol* 2016;11(1):93-102 [FREE Full text] [doi: [10.2217/fmb.15.99](https://doi.org/10.2217/fmb.15.99)] [Medline: [26323589](https://pubmed.ncbi.nlm.nih.gov/26323589/)]
25. McCullough AR, Rathbone J, Parekh S, Hoffmann TC, Del Mar CB. Not in my backyard: a systematic review of clinicians' knowledge and beliefs about antibiotic resistance. *J Antimicrob Chemother* 2015 Sep;70(9):2465-2473. [doi: [10.1093/jac/dkv164](https://doi.org/10.1093/jac/dkv164)] [Medline: [26093375](https://pubmed.ncbi.nlm.nih.gov/26093375/)]
26. Keizer J, Braakman-Jansen LM, Kampmeier S, Köck R, Al Naiemi N, Te Riet-Waring R, et al. Cross-border comparison of antimicrobial resistance (AMR) and AMR prevention measures: the healthcare workers' perspective. *Antimicrob Resist Infect Control* 2019 Jul 22;8(1):123. [doi: [10.1186/s13756-019-0577-4](https://doi.org/10.1186/s13756-019-0577-4)]
27. Colquhoun HL, Brehaut JC, Sales A, Ivers N, Grimshaw J, Michie S, et al. A systematic review of the use of theory in randomized controlled trials of audit and feedback. *Implement Sci* 2013 Jun 10;8:66 [FREE Full text] [doi: [10.1186/1748-5908-8-66](https://doi.org/10.1186/1748-5908-8-66)] [Medline: [23759034](https://pubmed.ncbi.nlm.nih.gov/23759034/)]
28. Davey P, Peden C, Charani E, Marwick C, Michie S. Time for action-Improving the design and reporting of behaviour change interventions for antimicrobial stewardship in hospitals: early findings from a systematic review. *Int J Antimicrob Agents* 2015 Mar;45(3):203-212 [FREE Full text] [doi: [10.1016/j.ijantimicag.2014.11.014](https://doi.org/10.1016/j.ijantimicag.2014.11.014)] [Medline: [25630430](https://pubmed.ncbi.nlm.nih.gov/25630430/)]
29. Lorencatto F, Charani E, Sevdalis N, Tarrant C, Davey P. Driving sustainable change in antimicrobial prescribing practice: how can social and behavioural sciences help? *J Antimicrob Chemother* 2018 Oct 01;73(10):2613-2624. [doi: [10.1093/jac/dky222](https://doi.org/10.1093/jac/dky222)] [Medline: [30020464](https://pubmed.ncbi.nlm.nih.gov/30020464/)]

30. Edwards R, Charani E, Sevdalis N, Alexandrou B, Sibley E, Mullett D, et al. Optimisation of infection prevention and control in acute health care by use of behaviour change: a systematic review. *Lancet Infectious Diseases* 2012 Apr;12(4):318-329. [doi: [10.1016/s1473-3099\(11\)70283-3](https://doi.org/10.1016/s1473-3099(11)70283-3)]
31. Hulscher ME, Grol RP, van der Meer JW. Antibiotic prescribing in hospitals: a social and behavioural scientific approach. *Lancet Infectious Diseases* 2010 Mar;10(3):167-175. [doi: [10.1016/s1473-3099\(10\)70027-x](https://doi.org/10.1016/s1473-3099(10)70027-x)]
32. Cruz-Martínez RR, Wentzel J, Asbjørnsen RA, Noort PD, van Niekerk JM, Sanderman R, et al. Supporting self-management of cardiovascular diseases through remote monitoring technologies: metaethnography review of frameworks, models, and theories used in research and development. *J Med Internet Res* 2020 May 21;22(5):e16157 [FREE Full text] [doi: [10.2196/16157](https://doi.org/10.2196/16157)] [Medline: [32436852](https://pubmed.ncbi.nlm.nih.gov/32436852/)]
33. van Buul LW, Sikkens JJ, van Agtmael MA, Kramer MH, van der Steen JT, Hertogh CM. Participatory action research in antimicrobial stewardship: a novel approach to improving antimicrobial prescribing in hospitals and long-term care facilities. *J Antimicrob Chemother* 2014 Jul;69(7):1734-1741. [doi: [10.1093/jac/dku068](https://doi.org/10.1093/jac/dku068)] [Medline: [24648505](https://pubmed.ncbi.nlm.nih.gov/24648505/)]
34. Gregor S, Hevner AR. Positioning and presenting design science research for maximum impact. *MISQ* 2013 Feb 2;37(2):337-355. [doi: [10.25300/misq/2013/37.2.01](https://doi.org/10.25300/misq/2013/37.2.01)]
35. Beerlage-de Jong N, Eikelenboom-Boskamp A, Voss A, Sanderman R, van Gemert-Pijnen JEW. Combining user-centered design with the persuasive systems design model: The development process of a web-based registration and monitoring system for healthcare-associated infections in nursing homes. *Int J Advances Life Sci* 2018;6(3):262-271 [FREE Full text]
36. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. *Commun Assoc Inf Syst* 2009;24. [doi: [10.17705/1cais.02428](https://doi.org/10.17705/1cais.02428)]
37. Wentzel J, van Drie-Pierik R, Nijdam L, Geesing J, Sanderman R, van Gemert-Pijnen JEW. Antibiotic information application offers nurses quick support. *Am J Infect Control* 2016 Jun 01;44(6):677-684. [doi: [10.1016/j.ajic.2015.12.038](https://doi.org/10.1016/j.ajic.2015.12.038)] [Medline: [26905792](https://pubmed.ncbi.nlm.nih.gov/26905792/)]
38. Leis JA. Advancing infection prevention and antimicrobial stewardship through improvement science. *BMJ Qual Saf* 2018 Feb 14;27(2):163-165. [doi: [10.1136/bmjqs-2017-006793](https://doi.org/10.1136/bmjqs-2017-006793)] [Medline: [28615344](https://pubmed.ncbi.nlm.nih.gov/28615344/)]
39. Birgand G, Johansson A, Szilagyi E, Lucet J. Overcoming the obstacles of implementing infection prevention and control guidelines. *Clin Microbiol Infect* 2015 Dec;21(12):1067-1071 [FREE Full text] [doi: [10.1016/j.cmi.2015.09.005](https://doi.org/10.1016/j.cmi.2015.09.005)] [Medline: [26369604](https://pubmed.ncbi.nlm.nih.gov/26369604/)]
40. Storr J, Twyman A, Zingg W, Damani N, Kilpatrick C, Reilly J, WHO Guidelines Development Group. Core components for effective infection prevention and control programmes: new WHO evidence-based recommendations. *Antimicrob Resist Infect Control* 2017;6:6 [FREE Full text] [doi: [10.1186/s13756-016-0149-9](https://doi.org/10.1186/s13756-016-0149-9)] [Medline: [28078082](https://pubmed.ncbi.nlm.nih.gov/28078082/)]
41. Davies P, Walker AE, Grimshaw JM. A systematic review of the use of theory in the design of guideline dissemination and implementation strategies and interpretation of the results of rigorous evaluations. *Implement Sci* 2010 Feb 09;5:14 [FREE Full text] [doi: [10.1186/1748-5908-5-14](https://doi.org/10.1186/1748-5908-5-14)] [Medline: [20181130](https://pubmed.ncbi.nlm.nih.gov/20181130/)]
42. Colquhoun HL, Carroll K, Eva KW, Grimshaw JM, Ivers N, Michie S, et al. Informing the research agenda for optimizing audit and feedback interventions: results of a prioritization exercise. *BMC Med Res Methodol* 2021 Jan 13;21(1):20 [FREE Full text] [doi: [10.1186/s12874-020-01195-5](https://doi.org/10.1186/s12874-020-01195-5)] [Medline: [33435873](https://pubmed.ncbi.nlm.nih.gov/33435873/)]
43. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. *J Med Internet Res* 2011 Dec 31;13(4):e126 [FREE Full text] [doi: [10.2196/jmir.1923](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]
44. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005 Feb;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
45. Peters MD, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. *Int J Evid Based Healthc* 2015 Sep;13(3):141-146. [doi: [10.1097/XEB.000000000000050](https://doi.org/10.1097/XEB.000000000000050)] [Medline: [26134548](https://pubmed.ncbi.nlm.nih.gov/26134548/)]
46. 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 Sep 04;169(7):467. [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)]
47. Eysenbach G. What is e-health? *J Med Internet Res* 2001;3(2):E20 [FREE Full text] [doi: [10.2196/jmir.3.2.e20](https://doi.org/10.2196/jmir.3.2.e20)] [Medline: [11720962](https://pubmed.ncbi.nlm.nih.gov/11720962/)]
48. Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014 Mar 07;348:g1687. [doi: [10.1136/bmj.g1687](https://doi.org/10.1136/bmj.g1687)] [Medline: [24609605](https://pubmed.ncbi.nlm.nih.gov/24609605/)]
49. 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](https://doi.org/10.1186/1748-5908-4-50)] [Medline: [19664226](https://pubmed.ncbi.nlm.nih.gov/19664226/)]
50. 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](https://doi.org/10.2196/jmir.8775)] [Medline: [29092808](https://pubmed.ncbi.nlm.nih.gov/29092808/)]
51. Tong A, Flemming K, McInnes E, Oliver S, Craig J. Enhancing transparency in reporting the synthesis of qualitative research: ENTREQ. *BMC Med Res Methodol* 2012 Nov 27;12(1):181 [FREE Full text] [doi: [10.1186/1471-2288-12-181](https://doi.org/10.1186/1471-2288-12-181)] [Medline: [23185978](https://pubmed.ncbi.nlm.nih.gov/23185978/)]

52. Boscart VM, Fernie GR, Lee JH, Jaglal SB. Using psychological theory to inform methods to optimize the implementation of a hand hygiene intervention. *Implement Sci* 2012 Aug 28;7(1):77 [FREE Full text] [doi: [10.1186/1748-5908-7-77](https://doi.org/10.1186/1748-5908-7-77)] [Medline: [22929925](https://pubmed.ncbi.nlm.nih.gov/22929925/)]
53. Conway LJ, Riley L, Saiman L, Cohen B, Alper P, Larson EL. Implementation and impact of an automated group monitoring and feedback system to promote hand hygiene among health care personnel. *Joint Commission J Qual Patient Safety* 2014 Sep;40(9):408-417. [doi: [10.1016/s1553-7250\(14\)40053-9](https://doi.org/10.1016/s1553-7250(14)40053-9)]
54. Edmisten C, Hall C, Kernizan L, Korwek K, Preston A, Rhoades E, et al. Implementing an electronic hand hygiene monitoring system: lessons learned from community hospitals. *Am J Infect Control* 2017 Aug 01;45(8):860-865. [doi: [10.1016/j.ajic.2017.03.033](https://doi.org/10.1016/j.ajic.2017.03.033)] [Medline: [28526308](https://pubmed.ncbi.nlm.nih.gov/28526308/)]
55. Hysong SJ, Kell HJ, Petersen LA, Campbell BA, Trautner BW. Theory-based and evidence-based design of audit and feedback programmes: examples from two clinical intervention studies. *BMJ Qual Saf* 2017 Apr 10;26(4):323-334. [doi: [10.1136/bmjqs-2015-004796](https://doi.org/10.1136/bmjqs-2015-004796)] [Medline: [27288054](https://pubmed.ncbi.nlm.nih.gov/27288054/)]
56. Smither JW, London M, Richmond KR. The relationship between leaders' personality and their reactions to and use of multisource feedback. *Group Organization Manag* 2016 Jul 26;30(2):181-210. [doi: [10.1177/1059601103254912](https://doi.org/10.1177/1059601103254912)]
57. James R, Upjohn L, Cotta M, Luu S, Marshall C, Buising K, et al. Measuring antimicrobial prescribing quality in Australian hospitals: development and evaluation of a national antimicrobial prescribing survey tool. *J Antimicrob Chemother* 2015 Feb 26;70(6):1912-1918. [doi: [10.1093/jac/dkv047](https://doi.org/10.1093/jac/dkv047)] [Medline: [25724986](https://pubmed.ncbi.nlm.nih.gov/25724986/)]
58. Jeanes A, Coen PG, Drey NS, Gould DJ. Moving beyond hand hygiene monitoring as a marker of infection prevention performance: development of a tailored infection control continuous quality improvement tool. *Am J Infect Control* 2020 Jan;48(1):68-76 [FREE Full text] [doi: [10.1016/j.ajic.2019.06.014](https://doi.org/10.1016/j.ajic.2019.06.014)] [Medline: [31358420](https://pubmed.ncbi.nlm.nih.gov/31358420/)]
59. Keizer J, Beerlage-de Jong N, Al Naiemi N, van Gemert-Pijnen JEW. Persuading from the start: participatory development of sustainable persuasive data-driven technologies in healthcare. In: *Persuasive Technology. Designing for Future Change*. Cham: Springer; 2020.
60. Marques R, Gregório J, Pinheiro F, Póvoa P, da Silva MM, Lapão LV. How can information systems provide support to nurses' hand hygiene performance? Using gamification and indoor location to improve hand hygiene awareness and reduce hospital infections. *BMC Med Inform Decis Mak* 2017 Jan 31;17(1):15 [FREE Full text] [doi: [10.1186/s12911-017-0410-z](https://doi.org/10.1186/s12911-017-0410-z)] [Medline: [28143613](https://pubmed.ncbi.nlm.nih.gov/28143613/)]
61. Pakyz AL, Moczygemba LR, VanderWielen LM, Edmond MB, Stevens MP, Kuzel AJ. Facilitators and barriers to implementing antimicrobial stewardship strategies: results from a qualitative study. *Am J Infect Control* 2014 Oct;42(10 Suppl):S257-S263. [doi: [10.1016/j.ajic.2014.04.023](https://doi.org/10.1016/j.ajic.2014.04.023)] [Medline: [25239719](https://pubmed.ncbi.nlm.nih.gov/25239719/)]
62. Parker V, Giles M, King J, Bantawa K. Barriers and facilitators to implementation of a multifaceted nurse-led intervention in acute care hospitals aimed at reducing indwelling urinary catheter use: a qualitative study. *J Clin Nurs* 2020 Aug 14;29(15-16):3042-3053. [doi: [10.1111/jocn.15337](https://doi.org/10.1111/jocn.15337)] [Medline: [32441867](https://pubmed.ncbi.nlm.nih.gov/32441867/)]
63. Patel SJ, Saiman L, Duchon JM, Evans D, Ferng Y, Larson E. Development of an antimicrobial stewardship intervention using a model of actionable feedback. *Interdiscip Perspect Infect Dis* 2012;2012:150367 [FREE Full text] [doi: [10.1155/2012/150367](https://doi.org/10.1155/2012/150367)] [Medline: [22500166](https://pubmed.ncbi.nlm.nih.gov/22500166/)]
64. Power M, Fogarty M, Madsen J, Fenton K, Stewart K, Brotherton A, et al. Learning from the design and development of the NHS Safety Thermometer. *Int J Qual Health Care* 2014 Jun 30;26(3):287-297 [FREE Full text] [doi: [10.1093/intqhc/mzu043](https://doi.org/10.1093/intqhc/mzu043)] [Medline: [24787136](https://pubmed.ncbi.nlm.nih.gov/24787136/)]
65. Nilsen P. Making sense of implementation theories, models and frameworks. *Implement Sci* 2015 Apr 21;10:53 [FREE Full text] [doi: [10.1186/s13012-015-0242-0](https://doi.org/10.1186/s13012-015-0242-0)] [Medline: [25895742](https://pubmed.ncbi.nlm.nih.gov/25895742/)]
66. Pulcini C, Binda F, Lamkang AS, Trett A, Charani E, Goff DA, et al. Developing core elements and checklist items for global hospital antimicrobial stewardship programmes: a consensus approach. *Clin Microbiol Infect* 2019 Jan;25(1):20-25 [FREE Full text] [doi: [10.1016/j.cmi.2018.03.033](https://doi.org/10.1016/j.cmi.2018.03.033)] [Medline: [29625170](https://pubmed.ncbi.nlm.nih.gov/29625170/)]
67. Guidelines on Core Components of Infection Prevention and Control Programmes at the National and Acute Health Care Facility Level. Genève, Switzerland: World Health Organization; 2017.
68. Kip H, Keizer J, da Silva MC, Beerlage-de Jong N, Köhle N, Kelders S. Methods for human-centered eHealth development: narrative scoping review. *J Med Internet Res* 2022 Jan 27;24(1):e31858 [FREE Full text] [doi: [10.2196/31858](https://doi.org/10.2196/31858)] [Medline: [35084359](https://pubmed.ncbi.nlm.nih.gov/35084359/)]
69. Birken SA, Powell BJ, Presseau J, Kirk MA, Lorencatto F, Gould NJ, et al. Combined use of the Consolidated Framework for Implementation Research (CFIR) and the Theoretical Domains Framework (TDF): a systematic review. *Implement Sci* 2017 Jan 05;12(1):2 [FREE Full text] [doi: [10.1186/s13012-016-0534-z](https://doi.org/10.1186/s13012-016-0534-z)] [Medline: [28057049](https://pubmed.ncbi.nlm.nih.gov/28057049/)]
70. 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](https://doi.org/10.1136/qshc.2004.011155)] [Medline: [15692000](https://pubmed.ncbi.nlm.nih.gov/15692000/)]
71. Tobi H, Kampen JK. Research design: the methodology for interdisciplinary research framework. *Qual Quant* 2018;52(3):1209-1225 [FREE Full text] [doi: [10.1007/s11135-017-0513-8](https://doi.org/10.1007/s11135-017-0513-8)] [Medline: [29674791](https://pubmed.ncbi.nlm.nih.gov/29674791/)]
72. CohenMiller AS, Pate E. A model for developing interdisciplinary research theoretical frameworks. *Qual Report* 2019 Jun 2;24(6):1211. [doi: [10.46743/2160-3715/2019.3558](https://doi.org/10.46743/2160-3715/2019.3558)]

73. Birken SA, Powell BJ, Shea CM, Haines ER, Alexis Kirk M, Leeman J, et al. Criteria for selecting implementation science theories and frameworks: results from an international survey. *Implement Sci* 2017 Oct 30;12(1):124 [FREE Full text] [doi: [10.1186/s13012-017-0656-y](https://doi.org/10.1186/s13012-017-0656-y)] [Medline: [29084566](https://pubmed.ncbi.nlm.nih.gov/29084566/)]
74. Yigitbasioğlu O, Velcu O. A review of dashboards in performance management: implications for design and research. *Int J Account Inf Syst* 2012 Mar;13(1):41-59. [doi: [10.1016/j.accinf.2011.08.002](https://doi.org/10.1016/j.accinf.2011.08.002)]
75. Kelders SM, Kip H, Greeff J. Psychometric evaluation of the TWente Engagement with Ehealth Technologies Scale (TWEETS): evaluation study. *J Med Internet Res* 2020 Oct 09;22(10):e17757 [FREE Full text] [doi: [10.2196/17757](https://doi.org/10.2196/17757)] [Medline: [33021487](https://pubmed.ncbi.nlm.nih.gov/33021487/)]
76. Gude WT, Brown B, van der Veer SN, Colquhoun HL, Ivers NM, Brehaut JC, et al. Clinical performance comparators in audit and feedback: a review of theory and evidence. *Implement Sci* 2019 Apr 24;14(1):39 [FREE Full text] [doi: [10.1186/s13012-019-0887-1](https://doi.org/10.1186/s13012-019-0887-1)] [Medline: [31014352](https://pubmed.ncbi.nlm.nih.gov/31014352/)]
77. Carey R, Connell L, Johnston M, Rothman A, de Bruin M, Kelly M, et al. Behavior change techniques and their mechanisms of action: a synthesis of links described in published intervention literature. *Ann Behav Med* 2019 Jul 17;53(8):693-707 [FREE Full text] [doi: [10.1093/abm/kay078](https://doi.org/10.1093/abm/kay078)] [Medline: [30304386](https://pubmed.ncbi.nlm.nih.gov/30304386/)]
78. Kip H, van Gemert-Pijnen JEW. Holistic development of eHealth technology. In: *eHealth Research, Theory and Development: A Multi-Disciplinary Approach*. London: Routledge; 2018:131-166.
79. Kip H, Beerlage-de Jong N, Wentzel J. The contextual inquiry. In: *eHealth Research, Theory and Development: A Multi-Disciplinary Approach*. London: Routledge; 2018:167-186.
80. Realising the potential of learning health systems. The Health Foundation. 2021. URL: <https://learninghealthcareproject.org/wp-content/uploads/2021/05/LHS2021report.pdf> [accessed 2022-03-01]
81. Mapping the Learning Health System: A Scoping Review of Current Evidence. A White Paper. New South Wales, Australia: Australian Institute of Health Innovation, Macquarie University; 2020.
82. Beerlage-de Jong N. *eHealth vs Infection: Participatory Development of Persuasive eHealth to Support Safe Care*. Enschede, The Netherlands: University of Twente; 2016.

Abbreviations

AF: audit and feedback

AMR: antimicrobial resistance

APM: antimicrobial resistance prevention measures

APM-AF: audit and feedback for antimicrobial resistance prevention measures

HCW: health care worker

HH: hand hygiene

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

TMF: theory, model, and framework

Edited by R Kukafka; submitted 24.09.21; peer-reviewed by T Hirose, E Laukka; comments to author 19.10.21; revised version received 01.12.21; accepted 03.12.21; published 11.03.22.

Please cite as:

Keizer J, Bente BE, Al Naiemi N, Van Gemert-Pijnen LJEWC, Beerlage-De Jong N

Improving the Development and Implementation of Audit and Feedback Systems to Support Health Care Workers in Limiting Antimicrobial Resistance in the Hospital: Scoping Review

J Med Internet Res 2022;24(3):e33531

URL: <https://www.jmir.org/2022/3/e33531>

doi: [10.2196/33531](https://doi.org/10.2196/33531)

PMID: [35275082](https://pubmed.ncbi.nlm.nih.gov/35275082/)

©Julia Keizer, Britt E Bente, Nashwan Al Naiemi, Lisette JEW Van Gemert-Pijnen, Nienke Beerlage-De Jong. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org/>), 11.03.2022. 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.

Review

Use of Mobile and Wearable Artificial Intelligence in Child and Adolescent Psychiatry: Scoping Review

Victoria Welch¹; Tom Joshua Wy², DO; Anna Ligezka³, MSc; Leslie C Hassett⁴, MLS; Paul E Croarkin², MSci, DO; Arjun P Athreya¹, MSc, PhD; Magdalena Romanowicz², MD

¹Department of Molecular Pharmacology and Experimental Therapeutics, Mayo Clinic, Rochester, MN, United States

²Department of Psychiatry and Psychology, Mayo Clinic, Rochester, MN, United States

³Department of Clinical Genomics, Mayo Clinic, Rochester, MN, United States

⁴Mayo Clinic Libraries, Mayo Clinic, Rochester, MN, United States

Corresponding Author:

Magdalena Romanowicz, MD

Department of Psychiatry and Psychology

Mayo Clinic

200 1st St SW

Rochester, MN, 55905

United States

Phone: 1 5072556782

Fax: 1 5072843933

Email: romanowicz.magdalena@mayo.edu

Abstract

Background: Mental health disorders are a leading cause of medical disabilities across an individual's lifespan. This burden is particularly substantial in children and adolescents because of challenges in diagnosis and the lack of precision medicine approaches. However, the widespread adoption of wearable devices (eg, smart watches) that are conducive for artificial intelligence applications to remotely diagnose and manage psychiatric disorders in children and adolescents is promising.

Objective: This study aims to conduct a scoping review to study, characterize, and identify areas of innovations with wearable devices that can augment current in-person physician assessments to individualize diagnosis and management of psychiatric disorders in child and adolescent psychiatry.

Methods: This scoping review used information from the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. A comprehensive search of several databases from 2011 to June 25, 2021, limited to the English language and excluding animal studies, was conducted. The databases included Ovid MEDLINE and Epub ahead of print, in-process and other nonindexed citations, and daily; Ovid Embase; Ovid Cochrane Central Register of Controlled Trials; Ovid Cochrane Database of Systematic Reviews; Web of Science; and Scopus.

Results: The initial search yielded 344 articles, from which 19 (5.5%) articles were left on the final source list for this scoping review. Articles were divided into three main groups as follows: studies with the main focus on autism spectrum disorder, attention-deficit/hyperactivity disorder, and internalizing disorders such as anxiety disorders. Most of the studies used either cardio-fitness chest straps with electrocardiogram sensors or wrist-worn biosensors, such as watches by Fitbit. Both allowed passive data collection of the physiological signals.

Conclusions: Our scoping review found a large heterogeneity of methods and findings in artificial intelligence studies in child psychiatry. Overall, the largest gap identified in this scoping review is the lack of randomized controlled trials, as most studies available were pilot studies and feasibility trials.

(*J Med Internet Res* 2022;24(3):e33560) doi:[10.2196/33560](https://doi.org/10.2196/33560)

KEYWORDS

mobile computing; artificial intelligence; wearable technologies; child psychiatry

Introduction

Background

The global burden of mental illnesses is daunting. In 2013, mental illnesses, such as major depression (2nd), anxiety disorders (7th), schizophrenia (11th), dysthymia (16th), and bipolar disorder (17th), were listed as some of the leading causes of years lived with disability [1]. Unipolar major depression and self-inflicted injuries were at number 2 and 14, respectively, as the top contributors to the global burden of disease in 2020 [2]. Despite significant advances in medical research, there is a distinct deficiency in the detection and treatment of psychiatric disorders, especially in children and adolescents [3]. The subsequent ramifications are evident. For instance, in 10% of completed suicides, the adolescent victims had no previously recorded psychiatric diagnoses [4]. With the increasing adoption of measurement-based practices in child and adolescent psychiatry, large volumes of data are being generated from clinical trials and routine practice. Such well-characterized data may prove to be fertile opportunities for innovations in data science and artificial intelligence (AI) to potentially address the shortcomings and subsequently improve the burden of disease in these populations.

Current approaches for the evaluation of psychiatric disorders predominantly rely on physician-patient history taking, collateral information, and patients' self-reported questionnaires rather than objective laboratory tests or neuroimaging biomarkers. As a result, contemporary psychiatric assessments are inaccurate and ineffective in providing a reliable and individualized assessment of symptoms [3]. With growing evidence of large data-driven approaches (eg, AI) to individualize diagnoses and treatment management of psychiatric disorders in adults [5-9], there are opportunities for such a paradigm in child and adolescent psychiatry. AI approaches simulate humans' ability to problem solve, plan, reason, and recognize patterns [3]. In these processes, computers *learn* the abilities by processing massive data sets through multilayered mathematical models (algorithms) and training methodologies (eg, cross-validation) improves the AI model's predictive confidence [10].

Broadly, the field of AI subsumes the methodological paradigms of computing science. First, machine learning refers to a programming approach in computer science in which the behavior of a program is not fully determined by an established code but can adapt its behavior (ie, learn) based on the data gathered [11]. Simple neural networks have been used in medicine since the early 1990s to interpret electrocardiograms (ECGs) [12], individualized predictions of antidepressant response, and diagnosis of myocardial infarction [13]. Second, deep learning is a particular variant of machine learning that is often modeled using artificial neural networks, which typically consist of interconnected nodes representing artificial neurons [11]. Deep learning has been used to design drugs, predict gene mutation expression, analyze histological examples, and read radiographic images [14,15]. Third, natural language processing (NLP) involves training computers to understand text and spoken languages or words in the way humans communicate [15]. NLP is well adopted in medicine, where it is used to extract structured

text (eg, diagnosis and treatment context) from unstructured text (eg, electronic health records). Finally, reinforcement learning is a field of computing where computers can be trained to make decisions based on past and current data and a given context to maximize long-term outcomes. For example, reinforcement learning has been used in tailoring treatments for epilepsy and sepsis [16]. These recent AI applications provide new possibilities for AI use in specialty medical practice, while projecting future utility in general medical practice [10].

Objectives

AI has been innovating and reshaping medicine but progress in psychiatry and child and adolescent psychiatry, in particular, is slow. Most previous research in psychiatry has focused on either NLP or the integration of various biomarkers to classify certain disorders such as heart conditions, epilepsy, and various types of cancer [11]. In addition, limited sustained collaboration among engineers, data scientists, and mental health providers has affected the slower adoption of these techniques in psychiatric practice in comparison with psychiatry and other medical specialties [10,17]. A review by Shatte et al [18] discussed the literature focused on adult mental health issues and machine learning applications. Our unique review aims to summarize the available research in child and adolescent psychiatry literature investigating machine learning technology and AI applications. The second aim is to characterize future opportunities in AI research in child and adolescent psychiatry.

The growing adoption of wearable technologies (eg, smart watches) not only helps passively collect large volumes of data but also opens the doors for using the data to enable remote diagnostic capabilities in child and adolescent psychiatry. Remote diagnosis and management of psychiatric diseases in children is crucial, given the shortage of trained mental health professionals.

Although wearable devices might support diagnosis and management in the future, they will not be able to replace health care professionals and their clinical observations. In this context, we sought to review studies that gathered data of child and adolescent patients with psychiatric disorders through practical means and used passively gathered data for various prediction mechanisms.

Methods

This scoping review used PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) information ([Multimedia Appendix 1](#)) [19] as a guide and was organized according to the steps outlined in this section.

Step 1: Developing a Research Question

We identified what was the extent of literature on the use of wearable devices in the form of AI in child and adolescent psychiatry. We aimed to gain insight into the following research objectives: (1) understand how wearable devices are being used in child psychiatry by exploring the types of devices that are being used, investigating how these devices are being used in child psychiatry (aid diagnosis, evaluate treatment efficacy, make algorithms to predict behavioral outcomes, etc), and researching which physiological signals are being measured by

these devices and (2) identify how the clinical knowledge of various pediatric psychiatric disorders has been expanded through the use of AI, specifically wearable devices.

Step 2: Literature Search

A comprehensive search of several databases was performed on June 25, 2021. The search was restricted from the year 2011 through the date the search was conducted. The results were limited to English language articles. Animal studies were excluded from this study. The databases searched were Ovid MEDLINE (≥ 1946) and Epub ahead of print, in-process and other nonindexed citations and daily (equivalent to PubMed); Ovid Embase (≥ 1988); Ovid Cochrane Central Register of Controlled Trials (≥ 1991); Ovid Cochrane Database of Systematic Reviews (≥ 2005); Web of Science Core Collection via Clarivate Analytics (≥ 1975); and Scopus via Elsevier (≥ 1788).

The search strategy was designed and conducted by an experienced librarian (LH), with input from the study investigators (APA and MR). Controlled vocabulary supplemented with keywords was used to search for studies. The actual strategy, listing all the search terms used and how they were combined, is available in [Multimedia Appendix 2](#).

Step 3: Study Selection

The study selection process was divided into two phases: (1) title and abstract screening and (2) full-text article screening. For the first phase, 2 reviewers (VW and JW) screened the articles and either excluded them or included them on the source list. The resulting source list was reviewed by 2 other reviewers (APA and MR). A full-text article review was then performed by 2 reviewers (VW and JW) and the final source list was created.

Step 4: Charting the Data

Most of the study data were extracted by a single researcher (VW). Another researcher (JW) helped complete 1 column and reviewed the table after it was completed. The following information was included in the table: year, sample population (size and demographics), psychiatric diagnosis, age range, wearable device used, and the measured physiological symptoms. The studies were divided based on the participants' main diagnosis: autism spectrum disorder (ASD),

attention-deficit/hyperactivity disorder (ADHD), and internalizing disorders (IDs).

Ethics and Dissemination

This proposed scoping review did not require ethics or institutional review board approval, as data were collected through the review of published peer-reviewed literature and gray literature. The results will be submitted for publication in an open-access peer-reviewed journal and presented at relevant medical and engineering conferences.

Results

Study Selection

The initial search yielded 344 articles. In all, 2 researchers (VW and TJW) completed the title and abstract screening process to narrow down this list. They reviewed papers independently, and disagreements were reviewed by 2 additional researchers (APA and MR). During this step, 89.2% (307/344) articles were omitted because they did not satisfy the inclusion criteria. The criteria for inclusion of studies for review in the first round of screening were as follows: the study must use a wearable device that passively tracked physiological variables in real time (objective measurements) and the device must be worn by a patient who had been diagnosed with one or more of the following psychiatric disorders: oppositional defiant disorder, conduct disorder, mood disorder (depression, anxiety, and bipolar disorder), ADHD, learning disability (dyslexia, etc), autism, or psychotic disorder. A full-text screening of the remaining 10.8% (37/344) articles was conducted by the same 2 researchers (VW and TJW). Given our focus on child and adolescent psychiatric illnesses, the criteria for study inclusion were accordingly altered to focus on child and adolescent psychiatric patients only (aged 0-18 years), wherein studies were written in English and published before January 1, 2021. Eligible designs included randomized controlled trials, nonrandomized experimental studies, cohort studies, and case-control studies. In addition, studies that focused on motor impairment were excluded. The PRISMA flow diagram for the study selection process is illustrated in [Figure 1](#). After the entire screening process was completed and reviewed by 2 additional researchers (APA and MR), 5.5% (19/344) articles were left on the final source list for this scoping review ([Table 1](#)).

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

PRISMA 2020 flow diagram for new systematic reviews which included searches of databases and registers only

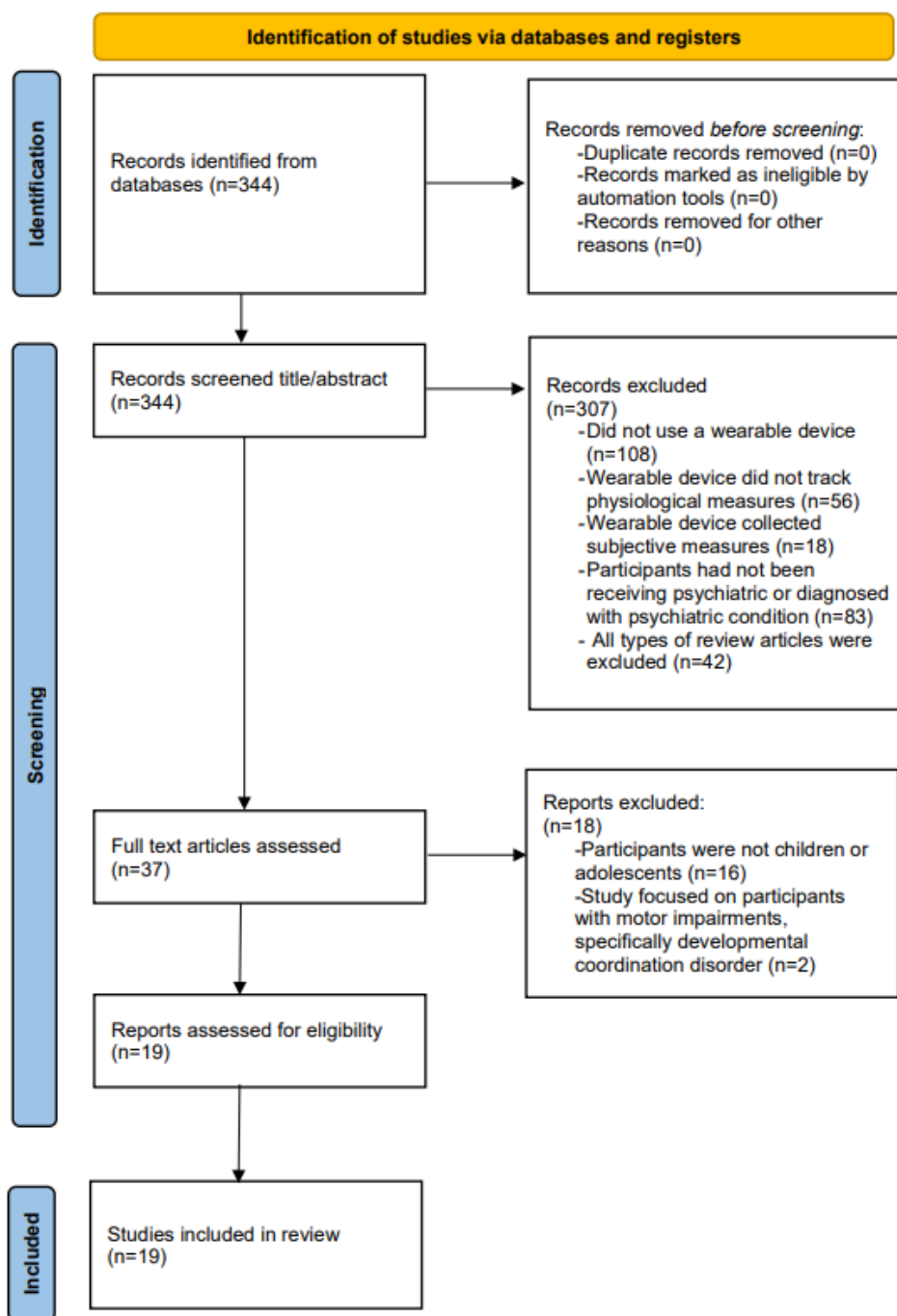


Table 1. Summary of studies on the use of wearable devices in child psychiatry.

Study	Year	Sample characteristics	Diagnosis	Age	Device	Measured physiological symptoms
Bilecci et al [20]	2018	40 participants: 29 males and 11 females; race not specified	ASD ^a	18-36 months	ECG ^b chest strap (Shimmer)	HR ^c , SDNN ^d , CV ^e , LF ^f , and HF ^g
Bilecci et al [21]	2016	5 participants: all males; race not specified	ASD	6-8 years	ECG chest strap and EEG ^h headset (EEG- Enobio wireless device)	QEEG ⁱ and HRV ^j
Di Palma et al [22]	2017	5 participants: all males; race not specified	ASD	6-8 years	ECG chest strap (based off Shimmer)	HR, RMSSD ^k , and RSA ^l
Faedda et al [23]	2016	155 participants: 97 males and 58 females; race not specified	BP ^m or ADHD ⁿ	5-18 years	ActiGraph belt (AMI ^o motionlogger)	Diurnal activity, sleep efficiency, and circadian regulation
Fioriello et al [24]	2020	24 participants: 18 males and 6 females; race not specified	ASD or LD ^p	30-72 months	ECG chest strap	HR
Gayet et al [25]	2014	not specified	ASD	Not specified	Accelerometer wrist strap (Affectiva Q Sensor), EEG headset (MindWave Mobile), ECG chest strap (Zephyr BioHarness), and mobile phone app (MyMedia)	Skin temperature and skin conductive data (accelerometer), EEG power spectrums (EEG), and HR, HRV, respiration, body temperature, and respiration (ECG)
Goodwin et al [26]	2019	20 participants: 75% male; 95% White; 90% non-Hispanic	ASD	6-17 years	Wrist-worn biosensor (Empatica E4)	HRV, EDA ^q , and motion-based activity (accelerometer)
Krupa et al [27]	2016	60 participants: sex and race not specified	ASD	3-12 years	Wrist-worn biosensor	HRV and EDA or GSR ^r
Kushki et al [28]	2015	24 participants: 17 males and 7 females; race not specified	ASD	Not specified	ECG chest strap (Shimmer)	RR ^s intervals
Leikauf et al [29]	2021	32 participants: 17 males and 15 females; race not specified	ADHD	8-17 years	Smart watch app (StopWatch)	Movement data (actigraphy via accelerometer)
Lin et al [30]	2020	30 participants: 11 males and 4 females with age-matched controls; race not specified	ADHD	5-9 years	Smart watch (Asus ZenWatch 3)	Angular velocity (gyroscope) and acceleration in axial direction (accelerometer)
McGinnis et al [31]	2021	63 participants: 57% female; 75% White, non-Latinx; 11% Asian or Pacific Islander; 11% African American; 3% biracial	IDs ^t	4-8 years	IMU ^u chest strap and headband (3-Space Sensor; YEI Technology)	Acceleration and angular velocity
McGinnis et al [32]	2019	63 participants: 57% female; 65% White; 82.5% in 2-parent households; 32% income >US \$100,000	IDs	3-8 years	IMU chest strap (3-Space Sensor; YEI Technology)	Acceleration and angular velocity
Min et al [33]	2011	4 participants: sex and race not specified	ASD	Not specified	Accelerometers worn on wrists, ankles, and upper body	Motion data (flapping, rocking, punching, and hitting)
Munoz-Organero et al [34]	2019	36 participants: 15 males and 3 females with non-matching controls; race not specified	ADHD	6-16 years	Accelerometers worn on wrists and ankles (Runscribe inertial sensors)	Acceleration and movement patterns
Ouyang et al [35]	2020	10 participants: sex and race not specified	ADHD	5-11 years	Accelerometer embedded in a smart watch	Linear motion

Study	Year	Sample characteristics	Diagnosis	Age	Device	Measured physiological symptoms
Pfeiffer et al [36]	2019	6 participants: sex not specified; 4 White; 2 Latin American or Hispanic	ASD	8-16 years	Wrist-worn biosensor (Empatica E4)	Skin conductance levels and NS-SCRs ^v (EDA data)
Redd et al [37]	2020	5 participants: sex and race not specified	IDs	8-12 years	Wrist-worn biosensor (Empatica E4)	HR, HRV, electrical property fluctuations in the skin (EDA data), motion (accelerometer), and peripheral skin temperature (infrared thermophile)
Wilson et al [38]	2021	5 participants: sex and race not specified	ASD	3-12 months	Ankle-worn biosensors (APDM Opal; APDM Wearable Technologies)	Motion complexity (accelerometer, gyroscope, and magnetometer)

^aASD: autism spectrum disorder.

^bECG: electrocardiogram.

^cHR: heart rate.

^dSDNN: SD of the averaged normal sinus RR intervals for 5-minute segments.

^eCV: time interval between 2 consecutive R waves.

^fLF: low frequency.

^gHF: high frequency.

^hEEG: electroencephalography.

ⁱQEEG: quantitative electroencephalography.

^jHRV: heart rate variability.

^kRMSSD: root-mean square of the successive normal sinus RR interval difference.

^lRSA: respiratory sinus arrhythmia (indicator of autonomic function).

^mBP: blood pressure.

ⁿADHD: attention-deficit/hyperactivity disorder.

^oAMI: acute myocardial infarction (motionlogger ActiGraph belt).

^pLD: learning disability.

^qEDA: electrodermal activity.

^rGSR: galvanic skin response.

^sRR interval, the time elapsed between 2 successive R waves of the QRS signal on the electrocardiogram.

^tID: internalizing disorder.

^uIMU: inertial measurement unit.

^vNS-SCR: nonspecific skin conductance response.

Introductory Information

The studies on the final source list were published between 2011 and 2021, and were spread across the world with most being conducted in the United States. In addition, the growth rate of research in this field increased substantially after 2018. The following sections provide a synopsis of each of the sources organized into the 3 categories previously outlined.

Studies Focused on ASD

Most studies (11/19, 58%) [20-22,24-28,33,36,38] focused on how the use of wearable devices could be used to aid in the treatment, behavioral prediction, and diagnosis of children with ASD. Although each study analyzed patients with ASD, the objectives and methods varied greatly among the studies. Several studies used ECG chest straps [20-22,24,25,28] for the most part to categorize the autonomic nervous system responses in patients with ASD during various tasks. For example, a study by Bilecci et al [20] used ECG strap during a joint attention stimuli in toddlers with ASD. The ECG chest strap measured

SD of the average normal sinus RR intervals (the time elapsed between 2 successive R waves of the QRS signal on the ECG) for 5-minute segments, heart rate (HR), CV (time interval between 2 consecutive R waves), low frequency (changes in sympathetic regulation), and high frequency (changes in parasympathetic regulation). The results showed that the SD of the average normal sinus RR intervals for 5-minute segments, CV, and low frequency values were significantly higher in the ASD group than in the control group at baseline. In addition, the CV was significantly higher in the ASD group during the joint attention task. These findings suggest that joint attention tasks coupled with wearable devices could potentially help physicians diagnose autism in toddlers.

A longitudinal study by Di Palma et al [22] mediated sociocognitive tasks through serious games allowed for the coding of child behavior. Children received treatment for 6 months while being monitored by an ECG wearable chest belt. The serious games consisted of joint attention tasks and imitation exercises and the ECG belt measured HR, respiratory sinus

arrhythmia (RSA; indicator of autonomic function), and root-mean square of the successive normal sinus RR interval difference. There was an increase in HR events during sociocognitive tasks. Researchers also found a correlation between detected physiological events and the level of involvement of the child during the task, along with a decrease in RSA and root-mean square of the successive normal sinus RR interval difference during activity which indicates proficient social interaction. Over time, patients displayed an increased percentage of physiological events associated with lower RSA during activity, which suggests improvement in cognitive engagement throughout the course of treatment. These predictive algorithms could be used at home by parents, at school by teachers, and in the clinic by therapists to create more individualized therapy plans.

Several studies used electroencephalography (EEG) in addition to ECG measures [21,25]. Bilecci et al [21] focused on obtaining quantitative EEG (pattern analysis of EEG) that was meant to determine treatment efficacy. Participants with ASD were monitored during a socioemotive interaction to implement a more individualized and effective treatment plan for these children. The results showed that all children yielded different measurements which emphasizes the importance of individualized therapy. The use of an AI device such as this would allow therapists to track a child's engagement so that they can tailor their therapy to the child's specific needs. Another study, in addition to measurement of EEG and ECG, also used an accelerometer wrist strap and created a mobile phone app called MyMedia and MySchedule [25]. The 6 main emotions that the app was designed to capture were happiness, sadness, fear, disgust, surprise, and anger. The sensors and facial recognition detected pupil dilation, skin conductance, HR or HR variability, blood pressure, concentration, and attention levels through a headset, watch, and chest strap. This method could provide a personalized way for autistic children and their caregivers to understand and manage their emotions.

Many of the studies used wrist-worn biosensors [26,27,36]. Some of the studies used accelerometers worn on various body parts such as wrists [25], ankles [38] as well as wrists, ankles, and upper body [33]. Using the Empatica E4 (Empatica, Inc) device study by Goodwin et al [26] investigated whether collecting and analyzing physiological and motion data from children with ASD during naturalistic observations could predict aggression. This wrist-worn sensor measured cardiovascular and electrodermal activity, along with detecting motion using an accelerometer. The results suggested that aggression to others can be predicted 1 minute before it occurs if biosensor data are collected for 3 minutes before the aggressive behavior. In this study, aggression was defined as hitting, kicking, biting, scratching, grabbing, pulling, pinching, or throwing objects at others. To make binary aggression predictions, a ridge-regularized logistic regression was used with the extracted time series features as input variables. This method had 84% average prediction accuracy and the average duration of aggressive episodes was 28 seconds.

Wilson et al [38] used wearable ankle sensors to diagnose ASD in children. Many believe that motor dysfunction may be predictive of ASD and the study used wearable ankle sensors

to track full-day motor activity in infants with a high familial risk for ASD. These sensors contained a 3D-accelerometer, a 3D-gyroscope, and a 3D-magnetometer. Leg movement data were collected when the participants were aged 3, 6, 9, and 12 months, and an autism diagnostic tool was used to evaluate each child at the ages of 18 and 36 months. On the basis of the movement data collected from the sensors, the researchers were able to construct a new measure of motion complexity defined in terms of the variability of the frequency components underlying the observed movements. The results of the study showed that high-risk infants with a later diagnosis of ASD showed lower motion complexity compared those who were not diagnosed later. In fact, there was a stronger correlation between motion complexity and ASD outcome relative to cognitive ability and adaptive skills, making this method a promising diagnostic tool.

A study by Krupa et al [27] also used a wearable wristband that measured the galvanic skin response and HR variability (indicators of the autonomic nervous system) for diagnostic purposes. The machine was also used to determine a child's current emotion. The results showed that the machine could differentiate children with ASD from normally developing children with 65% accuracy. In addition, it could differentiate neutral from emotion with 93.33% accuracy and happiness from involvement with 90% accuracy.

An intervention study by Pfeiffer et al [36] evaluated how well in-ear and overear headphones can decrease sympathetic activation in children with ASD and associated auditory hypersensitivity (hyperacusis) by measuring skin conductivity through electrodermal activity. Empatica E4 wristbands collected electrodermal activity data, as skin conductance is an indicator of stress or anxiety levels, and hyperacusis is associated with stress and anxiety. The results showed that in-ear and overear noise attenuating headphones led to a significant difference in both skin conductance levels and frequency of nonspecific conductance responses in subsequent phases of the study compared with the baseline measurements completed at the beginning of the study.

Studies Focused on ADHD

A small percentage of the studies (5/19, 26%) [23,29,30,34,35] examined how the use of wearable devices could be used to accurately diagnose ADHD and evaluate the treatment efficacy of various strategies used to treat children with ADHD. Although each of the studies in this category had different methods and objectives and collected different physiological measurements, all the devices used in these studies used an accelerometer [23,29,30,34,35].

In all, 11% (2/19) of studies used a smart watch application [29,30]. Leikauf et al [29] conducted a pilot study on the efficacy of StopWatch, a smart watch application designed to track movement and provide visual and haptic feedback regarding the movement data collected for patients with ADHD. As this application collected movement data via an accelerometer, this study focused specifically on the hyperactivity aspect of ADHD. In a similar study, researchers used data collected from the gyroscope and accelerometer placed in a smart watch to analyze the movements of children with ADHD [30]. They compared

the ADHD cohort to an age- and sex-matched control group to determine whether these types of measurements could be used to diagnose children with ADHD. After collecting data for 2 hours for 3 consecutive days in a naturalistic setting, the data suggested that children with ADHD had more variable and frequent movements than the controls. The Zero-Crossing Rate values across all 3 axes of the gyroscope were higher in the ADHD group; however, only the variance across the y-axis yielded a significant difference.

Munoz-Organero et al [34] used 4 triaxial accelerometers (placed on both wrists and ankles) to analyze movement patterns in normally developing children; patients with ADHD, who were medicated; and patients with ADHD, who were nonmedicated. They then used the data collected to propose a recurrent neural network to characterize the movement patterns of normally developing children that can be used to assess the similarity of new patients, which could potentially direct diagnosis. The results demonstrated that patients with ADHD, who were nonmedicated, showed higher differences medium intensity movements compared with normally developing patients whereas patients with ADHD, who were medicated, showed different behavior in their low intensity movements.

The final study in this category aimed to develop an objective method of evaluating the therapeutic effects of ADHD treatments by placing an accelerometer in their smart watch that recorded the movements of these patients [35]. The variance values of the accelerometer data before and after 1 month of using the medication methylphenidate were compared to determine the treatment efficacy of this drug. The Swanson, Nolan, and Pelham questionnaire (a subjective measure of treatment efficacy) was then compared with the accelerometer results, and the correlation between the 2 measurements was moderately strong. In addition, the variance values along the y- and z-axes of the accelerometers significantly decreased after 1 month of medication use, suggesting that the medication helped patients with ADHD.

A study by Faeda et al [23] used an ActiGraph belt (acute myocardial infarction motionlogger) and aimed to determine whether measures of activity, sleep, and circadian rhythm could be used to differentiate pediatric patients with bipolar disorder from pediatric patients with ADHD. They compared three different study groups: children with ADHD, children with bipolar disorder, children with ADHD and a comorbid depressive disorder and typically developing children. Each of these groups of children wore an ActiGraph belt for 3 to 5 days, which measured arousal, circadian rhythms, and sleep wakefulness cycles. The results showed that sleep duration and circadian strength measurements differed between children with ADHD and those with bipolar disorder. In addition, children with bipolar disorder had reduced measures of total sleep, reduced relative circadian amplitude, and increased nocturnal activity relative to the control group as well as both ADHD groups. This study suggests that wearable devices and AI may aid in the diagnosis of these overlapping pediatric disorders.

Studies Focused on IDs

A few studies (3/19, 16%) [31,32,37] investigated the identification children with IDs using wearable devices and

predicted their adverse behavioral outcomes through machine learning and AI.

McGinnis et al [31] aimed to develop a digital phenotype for childhood internalizing psychopathology based on data collected from a wearable inertial sensor. Data were recorded while the child completed three different tasks: the bubbles task (induced positivity), snake task (induced anxiety), and speech task (induced fear). The sensors were placed on both the head and waist of the child and acceleration and body movements were measured with an inertial measurement unit. They found that the children with IDs *burned out* quicker during the bubbles task and that it helped identify depressive, anxious, and trauma-related disorders. On the other hand, the snake task helped identify children with withdrawn, anxious, and depressive problems, oppositional problems, and specific phobias. Similarly, the speech task identified children with withdrawn, anxious, and depressive problems. After analyzing these results, they found that the phenotype from features that measured reward responsiveness could accurately detect children with underlying internalizing psychopathology with 75% accuracy.

Redd et al [37] investigated whether tracking physiological signals, such as HR, skin electrodermal activity, and skin temperature, could help predict a meltdown to facilitate earlier and more effective intervention. These measurements were recorded using a wrist-worn biosensor that incorporated blood volume pulsivity measurements, an electrodermal activity sensor to measure electrical fluctuations in the skin, a 3-axis accelerometer to measure overall motion and activity, and an infrared thermophile used to measure peripheral skin temperature. Parents also recorded their observations, and then the observations and physiological data were compared to create a predictive algorithm. The results showed that this model can accurately classify the behavioral states of children with 68% accuracy; however, only 4 meltdowns were recorded during the study which means that more data need to be collected and analyzed.

Discussion

Principal Findings

The primary outcome of this scoping review was a characterization of 19 studies of child and adolescent patients with ASD, ADHD and ID that gathered data through practical means and used passively gathered data for various prediction mechanisms.

From the included studies we were able to provide a descriptive analysis of currently available devices used to passively gather data in AI trials. Most of the studies used either ECG strap or wrist-worn biosensor. This is somewhat surprising as, for example, studies by Fioriello et al [24] and Kushki et al [28] could only collect data on HR via ECG chest strap, whereas many of the commercially available wrist watches can collect significantly more data [39].

The information obtained from this review can guide future trial development. Only 1 trial [23] had a high number of participants (N=155). Other trials [33] managed to enroll only 4 participants. We were not able to find any trials on children and adolescents

specifically with major depressive disorder or suicidality. Only 1 trial [23] compared children with ADHD and children with bipolar disorder; however, no studies specifically addressed children with bipolar disorder or psychosis. There were also no trials that gathered data during inpatient hospitalization or during evidence-based outpatient treatment that children would typically receive. No studies obtained data from parents as well during the trials focusing only on children.

All the studies summarized focused on detecting behavioral changes in patients. This leaves opportunities to innovate with the data to actuate the detected behaviors using reinforcement learning approaches. Particularly, if a state of behavior is predicted before it visibly manifests, there is an opportunity to alert responsible adults nearby to intervene using proven interventions. Such timely interventions may prove to be a positive reinforcement for children, who otherwise are penalized for unacceptable behavior. In older patients (ie, adolescents vs young children), physiology-based triggers (eg, based on HR variability) with an action message (eg, *performing breathing exercises for 30 seconds*) may be helpful in forcing a positive change in behavior or mood.

Feature extraction and feature engineering are the key aspects of machine learning and AI efforts using wearable data [40]. Given the heterogeneity in the types of devices and clinical questions being researched, approaches to extract signals from wearable data widely differed among the literature reviewed in this work. Feature extraction methods (eg, principal-component analysis, locally linear embedding, and autoencoding) will facilitate the identification of wearable-derived measures associated with clinical diagnosis or outcomes. Feature engineering uses wearable-derived measures to develop features for downstream analyses using domain-guided knowledge (eg, deriving motion features from raw accelerometer data). Determining the feature extraction or engineering approach is dependent on the nature of the clinical question being posed and types of measurements derived from wearables.

As AI in child and adolescent psychiatry is an emerging field, there are obvious gaps that will continue to be explored. Most of the papers focused on using the technology for diagnostic purposes, specifically ASD and ADHD. We believe that these emerging technologies can be used in other diagnostic categories such as mood, anxiety, and psychosis. There is also much room to innovate in the realm of intervention, treatment, and public health.

Limitations

Our study had several limitations. We did not conduct a systematic review or prospectively register a protocol. This was expected to be a nascent area; hence, a scoping review was most appropriate. Another limitation is the potential that we have missed important original literature on the use of mobile and wearable AI in child psychiatry. This was mitigated by an extensive search of multiple databases, searching references of

included articles, and ensuring duplicate review of all the abstracts and full-text. We chose not to include nonoriginal or non-peer-reviewed research and non-English articles. This might have led to us missing key conclusions drawn from this research. We also did not examine each of the machine learning technologies in detail but rather resorted to a brief description of the methods used in each of the studies.

Conclusions

This scoping review provides a comprehensive assessment of the literature on the use of mobile and wearable AI in children with ASD, ADHD, and ID. Our scoping review found large heterogeneity of methods and findings in AI studies in child psychiatry. Overall, the largest gaps identified in this scoping review are the lack of randomized controlled trials, as most studies available were pilot trials. The definition of digital biomarkers used in these studies seems to be very wide. The studies included in the review had a small number of participants. Nevertheless, our scoping review identified several key strengths across the disorders considered in this study. First, wearable technologies comprising multiple sensors (eg, HR, sleep, and accelerometers) demonstrate promise in the diagnosis and prediction of aspects of disorders spanning child and adolescent psychiatry. Second, analytic solutions using data from wearables and expert annotations of child behavior can predict the onset of behavioral changes relevant to psychiatric disorders. Hence, given the growing ubiquity of wearables across the age span (children to parents, guardians, or teachers), our review strongly suggests the incorporation of wearables in child and adolescent psychiatry research. Such integration would be pivotal in facilitating remote monitoring and remote psychiatric services, which will likely help reduce disparities in mental health care access because of a shortage of child and adolescent psychiatrists. From a research perspective, the interaction of wearable biomarkers with conventional biomarkers (eg, genomics, metabolomics, neuroimaging, EEG, and environmental exposures) in the context of the diagnosis, treatment, and management of psychiatric disorders is yet to be pursued in large studies. Finally, the integration of wearables in child and adolescent psychiatry research should extend beyond controlled research settings to allow the extraction of the benefits of AI approaches. Naturalistic studies should look to collect annotations of children's behavior as observed in their daily life at home, day care, or school and expand the involvement of relevant stakeholders in studies, wherein not only are parents annotating behaviors but also teachers, social workers, and counselors. Such a collection of annotated data from a real-world environment where children and adolescents develop will also provide opportunities to innovate with AI approaches such as reinforcement learning. Future directions should focus specifically on enrolling larger number of more diverse groups of patients. Future research should also focus on assessing which tools, mobile and wearable, are most efficient in collecting the most reliable data in various patient populations, as the primary outcome of interest.

Acknowledgments

This study was partially supported by the Mayo Clinic Summer Undergraduate Research Fellowship, National Science Foundation under grant 2041339; National Institutes of Health under grants R01 MH113700 and R01 MH124655; the Mayo Clinic Center for Individualized Medicine, and the Mayo Clinic Center for Clinical and Translational Science. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or National Institutes of Health.

Conflicts of Interest

PC has received research support from Pfizer, Inc and equipment support from Neuronetics, Inc and MagVenture, Inc. He has received grant in kind support from AssureRX for supplies and genotyping. He has been the primary investigator for a multicenter study funded by Neuronetics, Inc and a site primary investigator for a study funded by NeoSync, Inc. He has served as a paid consultant for Engrail Therapeutics, Myriad Neuroscience, Procter and Gamble Company, and Sunovion.

Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) scoping review checklist.

[[DOCX File , 109 KB - jmir_v24i3e33560_app1.docx](#)]

Multimedia Appendix 2

Actual search strategies used in the study.

[[DOCX File , 20 KB - jmir_v24i3e33560_app2.docx](#)]

References

1. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet* 2015;386(9995):743-800 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(15\)60692-4](https://doi.org/10.1016/S0140-6736(15)60692-4)] [Medline: [26063472](https://pubmed.ncbi.nlm.nih.gov/26063472/)]
2. Lopez AD, Murray CC. The global burden of disease, 1990-2020. *Nat Med* 1998;4(11):1241-1243. [doi: [10.1038/3218](https://doi.org/10.1038/3218)] [Medline: [9809543](https://pubmed.ncbi.nlm.nih.gov/9809543/)]
3. Fakhoury M. Artificial intelligence in psychiatry. *Adv Exp Med Biol* 2019;1192:119-125. [doi: [10.1007/978-981-32-9721-0_6](https://doi.org/10.1007/978-981-32-9721-0_6)] [Medline: [31705492](https://pubmed.ncbi.nlm.nih.gov/31705492/)]
4. Bridge JA, Goldstein TR, Brent DA. Adolescent suicide and suicidal behavior. *J Child Psychol Psychiatry* 2006;47(3-4):372-394. [doi: [10.1111/j.1469-7610.2006.01615.x](https://doi.org/10.1111/j.1469-7610.2006.01615.x)] [Medline: [16492264](https://pubmed.ncbi.nlm.nih.gov/16492264/)]
5. Athreya AP, Neavin D, Carrillo-Roa T, Skime M, Biernacka J, Frye MA, et al. Pharmacogenomics-driven prediction of antidepressant treatment outcomes: a machine-learning approach with multi-trial replication. *Clin Pharmacol Ther* 2019;106(4):855-865 [[FREE Full text](#)] [doi: [10.1002/cpt.1482](https://doi.org/10.1002/cpt.1482)] [Medline: [31012492](https://pubmed.ncbi.nlm.nih.gov/31012492/)]
6. Athreya AP, Iyer R, Wang L, Weinshilboum RM, Bobo WV. Integration of machine learning and pharmacogenomic biomarkers for predicting response to antidepressant treatment: can computational intelligence be used to augment clinical assessments? *Pharmacogenomics* 2019;20(14):983-988 [[FREE Full text](#)] [doi: [10.2217/pgs-2019-0119](https://doi.org/10.2217/pgs-2019-0119)] [Medline: [31559920](https://pubmed.ncbi.nlm.nih.gov/31559920/)]
7. Athreya AP, Brückl T, Binder EB, John Rush A, Biernacka J, Frye MA, et al. Prediction of short-term antidepressant response using probabilistic graphical models with replication across multiple drugs and treatment settings. *Neuropsychopharmacology* 2021;46(7):1272-1282 [[FREE Full text](#)] [doi: [10.1038/s41386-020-00943-x](https://doi.org/10.1038/s41386-020-00943-x)] [Medline: [33452433](https://pubmed.ncbi.nlm.nih.gov/33452433/)]
8. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. *Lancet Psychiatry* 2016;3(3):243-250. [doi: [10.1016/S2215-0366\(15\)00471-X](https://doi.org/10.1016/S2215-0366(15)00471-X)] [Medline: [26803397](https://pubmed.ncbi.nlm.nih.gov/26803397/)]
9. Chekroud AM, Gueorguieva R, Krumholz HM, Trivedi MH, Krystal JH, McCarthy G. Reevaluating the efficacy and predictability of antidepressant treatments: a symptom clustering approach. *JAMA Psychiatry* 2017;74(4):370-378 [[FREE Full text](#)] [doi: [10.1001/jamapsychiatry.2017.0025](https://doi.org/10.1001/jamapsychiatry.2017.0025)] [Medline: [28241180](https://pubmed.ncbi.nlm.nih.gov/28241180/)]
10. Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer? *Am J Med* 2018;131(2):129-133. [doi: [10.1016/j.amjmed.2017.10.035](https://doi.org/10.1016/j.amjmed.2017.10.035)] [Medline: [29126825](https://pubmed.ncbi.nlm.nih.gov/29126825/)]
11. Brunn M, Diefenbacher A, Courtet P, Genieys W. The future is knocking: how artificial intelligence will fundamentally change psychiatry. *Acad Psychiatry* 2020;44(4):461-466. [doi: [10.1007/s40596-020-01243-8](https://doi.org/10.1007/s40596-020-01243-8)] [Medline: [32424706](https://pubmed.ncbi.nlm.nih.gov/32424706/)]
12. Willems JL, Abreu-Lima C, Arnaud P, van Bommel JH, Brohet C, Degani R, et al. The diagnostic performance of computer programs for the interpretation of electrocardiograms. *N Engl J Med* 1991;325(25):1767-1773. [doi: [10.1056/NEJM199112193252503](https://doi.org/10.1056/NEJM199112193252503)] [Medline: [1834940](https://pubmed.ncbi.nlm.nih.gov/1834940/)]
13. Baxt WG. Use of an artificial neural network for the diagnosis of myocardial infarction. *Ann Intern Med* 1991;115(11):843-848. [doi: [10.7326/0003-4819-115-11-843](https://doi.org/10.7326/0003-4819-115-11-843)] [Medline: [1952470](https://pubmed.ncbi.nlm.nih.gov/1952470/)]

14. Leung MK, Xiong HY, Lee LJ, Frey BJ. Deep learning of the tissue-regulated splicing code. *Bioinformatics* 2014;30(12):i121-i129 [FREE Full text] [doi: [10.1093/bioinformatics/btu277](https://doi.org/10.1093/bioinformatics/btu277)] [Medline: [24931975](https://pubmed.ncbi.nlm.nih.gov/24931975/)]
15. Chowdhury GG. Natural language processing. *Ann Rev Info Sci Tech* 2003;37(1):51-89. [doi: [10.1002/aris.1440370103](https://doi.org/10.1002/aris.1440370103)]
16. Gottesman O, Johansson F, Komorowski M, Faisal A, Sontag D, Doshi-Velez F, et al. Guidelines for reinforcement learning in healthcare. *Nat Med* 2019;25(1):16-18. [doi: [10.1038/s41591-018-0310-5](https://doi.org/10.1038/s41591-018-0310-5)] [Medline: [30617332](https://pubmed.ncbi.nlm.nih.gov/30617332/)]
17. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, et al. Artificial intelligence for mental health and mental illnesses: an overview. *Curr Psychiatry Rep* 2019;21(11):116 [FREE Full text] [doi: [10.1007/s11920-019-1094-0](https://doi.org/10.1007/s11920-019-1094-0)] [Medline: [31701320](https://pubmed.ncbi.nlm.nih.gov/31701320/)]
18. Shatte AB, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. *Psychol Med* 2019;49(9):1426-1448. [doi: [10.1017/S0033291719000151](https://doi.org/10.1017/S0033291719000151)] [Medline: [30744717](https://pubmed.ncbi.nlm.nih.gov/30744717/)]
19. 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;169(7):467-473 [FREE Full text] [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
20. Billeci L, Tonacci A, Narzisi A, Manigrasso Z, Varanini M, Fulceri F, et al. Heart rate variability during a joint attention task in toddlers with autism spectrum disorders. *Front Physiol* 2018;9:467 [FREE Full text] [doi: [10.3389/fphys.2018.00467](https://doi.org/10.3389/fphys.2018.00467)] [Medline: [29765335](https://pubmed.ncbi.nlm.nih.gov/29765335/)]
21. Billeci L, Tonacci A, Tartarisco G, Narzisi A, Di Palma S, Corda D, Michelangelo Study Group. An integrated approach for the monitoring of brain and autonomic response of children with autism spectrum disorders during treatment by wearable technologies. *Front Neurosci* 2016;10:276 [FREE Full text] [doi: [10.3389/fnins.2016.00276](https://doi.org/10.3389/fnins.2016.00276)] [Medline: [27445652](https://pubmed.ncbi.nlm.nih.gov/27445652/)]
22. Di Palma S, Tonacci A, Narzisi A, Domenici C, Pioggia G, Muratori F, MICHELANGELO study group. Monitoring of autonomic response to sociocognitive tasks during treatment in children with autism spectrum disorders by wearable technologies: a feasibility study. *Comput Biol Med* 2017;85:143-152. [doi: [10.1016/j.combiomed.2016.04.001](https://doi.org/10.1016/j.combiomed.2016.04.001)] [Medline: [27080078](https://pubmed.ncbi.nlm.nih.gov/27080078/)]
23. Faedda GL, Ohashi K, Hernandez M, McGreenery CE, Grant MC, Baroni A, et al. Actigraph measures discriminate pediatric bipolar disorder from attention-deficit/hyperactivity disorder and typically developing controls. *J Child Psychol Psychiatry* 2016;57(6):706-716 [FREE Full text] [doi: [10.1111/jcpp.12520](https://doi.org/10.1111/jcpp.12520)] [Medline: [26799153](https://pubmed.ncbi.nlm.nih.gov/26799153/)]
24. Fioriello F, Maugeri A, D'Alvia L, Pittella E, Piuze E, Rizzuto E, et al. A wearable heart rate measurement device for children with autism spectrum disorder. *Sci Rep* 2020;10(1):18659 [FREE Full text] [doi: [10.1038/s41598-020-75768-1](https://doi.org/10.1038/s41598-020-75768-1)] [Medline: [33122808](https://pubmed.ncbi.nlm.nih.gov/33122808/)]
25. Gay V, Leijdekkers P. Design of emotion-aware mobile apps for autistic children. *Health Technol* 2014;4(1):21-26. [doi: [10.1007/s12553-013-0066-3](https://doi.org/10.1007/s12553-013-0066-3)]
26. Goodwin MS, Mazefsky CA, Ioannidis S, Erdogmus D, Siegel M. Predicting aggression to others in youth with autism using a wearable biosensor. *Autism Res* 2019;12(8):1286-1296 [FREE Full text] [doi: [10.1002/aur.2151](https://doi.org/10.1002/aur.2151)] [Medline: [31225952](https://pubmed.ncbi.nlm.nih.gov/31225952/)]
27. Krupa N, Anantharam K, Sanker M, Datta S, Sagar JV. Recognition of emotions in autistic children using physiological signals. *Health Technol* 2016;6(2):137-147. [doi: [10.1007/s12553-016-0129-3](https://doi.org/10.1007/s12553-016-0129-3)]
28. Kushki A, Khan A, Brian J, Anagnostou E. A Kalman filtering framework for physiological detection of anxiety-related arousal in children with autism spectrum disorder. *IEEE Trans Biomed Eng* 2015;62(3):990-1000. [doi: [10.1109/tbme.2014.2377555](https://doi.org/10.1109/tbme.2014.2377555)]
29. Leikauf JE, Correa C, Bueno AN, Sempere VP, Williams LM. StopWatch: pilot study for an Apple Watch application for youth with ADHD. *Digit Health* 2021;7:20552076211001215 [FREE Full text] [doi: [10.1177/20552076211001215](https://doi.org/10.1177/20552076211001215)] [Medline: [33868703](https://pubmed.ncbi.nlm.nih.gov/33868703/)]
30. Lin LC, Ouyang CS, Chiang CT, Wu RC, Yang RC. Quantitative analysis of movements in children with attention-deficit hyperactivity disorder using a smart watch at school. *Appl Sci* 2020;10(12):4116. [doi: [10.3390/app10124116](https://doi.org/10.3390/app10124116)]
31. McGinnis EW, Scism J, Hruschak J, Muzik M, Rosenblum KL, Fitzgerald K, et al. Digital phenotype for childhood internalizing disorders: less positive play and promise for a brief assessment battery. *IEEE J Biomed Health Inform* 2021;25(8):3176-3184. [doi: [10.1109/JBHI.2021.3053846](https://doi.org/10.1109/JBHI.2021.3053846)] [Medline: [33481724](https://pubmed.ncbi.nlm.nih.gov/33481724/)]
32. McGinnis RS, McGinnis EW, Hruschak J, Lopez-Duran NL, Fitzgerald K, Rosenblum KL, et al. Rapid detection of internalizing diagnosis in young children enabled by wearable sensors and machine learning. *PLoS One* 2019;14(1):e0210267 [FREE Full text] [doi: [10.1371/journal.pone.0210267](https://doi.org/10.1371/journal.pone.0210267)] [Medline: [30650109](https://pubmed.ncbi.nlm.nih.gov/30650109/)]
33. Min CH, Tewfik AH. Semi-supervised event detection using higher order statistics for multidimensional time series accelerometer data. *Annu Int Conf IEEE Eng Med Biol Soc* 2011;2011:365-368. [doi: [10.1109/IEMBS.2011.6090119](https://doi.org/10.1109/IEMBS.2011.6090119)] [Medline: [22254324](https://pubmed.ncbi.nlm.nih.gov/22254324/)]
34. Muñoz-Organero M, Powell L, Heller B, Harpin V, Parker J. Using recurrent neural networks to compare movement patterns in ADHD and normally developing children based on acceleration signals from the wrist and ankle. *Sensors (Basel)* 2019;19(13):2935 [FREE Full text] [doi: [10.3390/s19132935](https://doi.org/10.3390/s19132935)] [Medline: [31277297](https://pubmed.ncbi.nlm.nih.gov/31277297/)]
35. Ouyang CS, Yang RC, Chiang CT, Wu RC, Lin LC. Objective evaluation of therapeutic effects of ADHD medication using a smart watch: a pilot study. *Appl Sci* 2020;10(17):5946. [doi: [10.3390/app10175946](https://doi.org/10.3390/app10175946)]

36. Pfeiffer B, Stein Duker L, Murphy A, Shui C. Effectiveness of noise-attenuating headphones on physiological responses for children with autism spectrum disorders. *Front Integr Neurosci* 2019;13:65 [FREE Full text] [doi: [10.3389/fnint.2019.00065](https://doi.org/10.3389/fnint.2019.00065)] [Medline: [31798424](https://pubmed.ncbi.nlm.nih.gov/31798424/)]
37. Redd C, Silvera-Tawil D, Hopp D, Zandberg D, Martiniuk A, Dietrich C, et al. Physiological signal monitoring for identification of emotional dysregulation in children. *Annu Int Conf IEEE Eng Med Biol Soc* 2020;2020:4273-4277. [doi: [10.1109/EMBC44109.2020.9176506](https://doi.org/10.1109/EMBC44109.2020.9176506)] [Medline: [33018940](https://pubmed.ncbi.nlm.nih.gov/33018940/)]
38. Wilson RB, Vangala S, Elashoff D, Safari T, Smith BA. Using wearable sensor technology to measure motion complexity in infants at high familial risk for autism spectrum disorder. *Sensors (Basel)* 2021;21(2):616 [FREE Full text] [doi: [10.3390/s21020616](https://doi.org/10.3390/s21020616)] [Medline: [33477359](https://pubmed.ncbi.nlm.nih.gov/33477359/)]
39. Gaudet J, Gallant F, Bélanger M. A bit of fit: minimalist intervention in adolescents based on a physical activity tracker. *JMIR Mhealth Uhealth* 2017;5(7):e92 [FREE Full text] [doi: [10.2196/mhealth.7647](https://doi.org/10.2196/mhealth.7647)] [Medline: [28684384](https://pubmed.ncbi.nlm.nih.gov/28684384/)]
40. Hulzink J, Konijnenburg M, Ashouei M, Breeschoten A, Berset T, Huisken J, et al. An ultra low energy biomedical signal processing system operating at near-threshold. *IEEE Trans Biomed Circuits Syst* 2011;5(6):546-554. [doi: [10.1109/TBCAS.2011.2176726](https://doi.org/10.1109/TBCAS.2011.2176726)] [Medline: [23852552](https://pubmed.ncbi.nlm.nih.gov/23852552/)]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder

AI: artificial intelligence

ASD: autism spectrum disorder

ECG: electrocardiogram

EEG: electroencephalography

HR: heart rate

ID: internalizing disorder

NLP: natural language processing

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

RSA: respiratory sinus arrhythmia

Edited by A Mavragani; submitted 13.09.21; peer-reviewed by S Pang, K Bul; comments to author 25.11.21; revised version received 13.01.22; accepted 26.01.22; published 14.03.22.

Please cite as:

Welch V, Wy TJ, Ligezka A, Hassett LC, Croarkin PE, Athreya AP, Romanowicz M

Use of Mobile and Wearable Artificial Intelligence in Child and Adolescent Psychiatry: Scoping Review

J Med Internet Res 2022;24(3):e33560

URL: <https://www.jmir.org/2022/3/e33560>

doi: [10.2196/33560](https://doi.org/10.2196/33560)

PMID: [35285812](https://pubmed.ncbi.nlm.nih.gov/35285812/)

©Victoria Welch, Tom Joshua Wy, Anna Ligezka, Leslie C Hassett, Paul E Croarkin, Arjun P Athreya, Magdalena Romanowicz. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 14.03.2022. 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.

Review

Workarounds in Electronic Health Record Systems and the Revised Sociotechnical Electronic Health Record Workaround Analysis Framework: Scoping Review

Vincent Blijleven¹, PhD; Florian Hoxha², BSc; Monique Jaspers², Prof Dr

¹Center for Marketing & Supply Chain Management, Nyenrode Business Universiteit, Breukelen, Netherlands

²Center for Human Factors Engineering of Health Information Technology, Amsterdam Public Health Research Institute, Amsterdam UMC, Amsterdam, Netherlands

Corresponding Author:

Vincent Blijleven, PhD

Center for Marketing & Supply Chain Management

Nyenrode Business Universiteit

Straatweg 25

Breukelen, 3621 BG

Netherlands

Phone: 31 630023248

Email: vincentblijleven@gmail.com

Abstract

Background: Electronic health record (EHR) system users devise workarounds to cope with mismatches between workflows designed in the EHR and preferred workflows in practice. Although workarounds appear beneficial at first sight, they frequently jeopardize patient safety, the quality of care, and the efficiency of care.

Objective: This review aims to aid in identifying, analyzing, and resolving EHR workarounds; the Sociotechnical EHR Workaround Analysis (SEWA) framework was published in 2019. Although the framework was based on a large case study, the framework still required theoretical validation, refinement, and enrichment.

Methods: A scoping literature review was performed on studies related to EHR workarounds published between 2010 and 2021 in the MEDLINE, Embase, CINAHL, Cochrane, or IEEE databases. A total of 737 studies were retrieved, of which 62 (8.4%) were included in the final analysis. Using an analytic framework, the included studies were investigated to uncover the rationales that EHR users have for workarounds, attributes characterizing workarounds, possible scopes, and types of perceived impacts of workarounds.

Results: The SEWA framework was theoretically validated and extended based on the scoping review. Extensive support for the pre-existing rationales, attributes, possible scopes, and types of impact was found in the included studies. Moreover, 7 new rationales, 4 new attributes, and 3 new types of impact were incorporated. Similarly, the descriptions of multiple pre-existing rationales for workarounds were refined to describe each rationale more accurately.

Conclusions: SEWA is now grounded in the existing body of peer-reviewed empirical evidence on EHR workarounds and, as such, provides a theoretically validated and more complete synthesis of EHR workaround rationales, attributes, possible scopes, and types of impact. The revised SEWA framework can aid researchers and practitioners in a wider range of health care settings to identify, analyze, and resolve workarounds. This will improve user-centered EHR design and redesign, ultimately leading to improved patient safety, quality of care, and efficiency of care.

(*J Med Internet Res* 2022;24(3):e33046) doi:[10.2196/33046](https://doi.org/10.2196/33046)

KEYWORDS

electronic health records; electronic medical records; framework; patient safety; unintended consequences; usability; workarounds; workflow

Introduction

Electronic health record (EHR) systems are the backbone of modern health care organizations. This is in pursuit of promising gains in patient safety, quality of care, efficiency, and control of spiraling costs by enabling value-based reimbursements. However, realizing these expected benefits is far from a given value. Over the years, an overwhelming number of studies have reported that EHRs have led to a multitude of unintended consequences. Examples include potential patient harm resulting from bad EHR usability [1,2]; increased odds of burnout of health care professionals [3,4]; physicians experiencing stress [5]; users spending an equal amount of time on *desktop medicine* as they spend on having face-to-face interaction with patients [6,7]; extensive *copy and paste* practices of patient notes leading to note bloating, internal inconsistencies, and errors [8]; and the unavailability of complete clinical information at the point of care [9].

Many causes of unintended consequences of EHR use can be traced back to discrepancies between the behavior, intentions, and expectations of EHR users and the workflows dictated by EHRs [10-15]. When EHR users experience workflow mismatches, they often create workarounds [16]. Workarounds are practices that handle exceptions to normal workflow [17] and do not follow the rules, assumptions, workflow regulations, or intentions of systems designers [18]. Although workarounds allow EHR users to proceed in accomplishing tasks in their preferred way (with or without the EHR), research shows that workarounds frequently jeopardize the safety, quality, and efficiency of care [19]. Given their common adverse effects, workarounds are valuable points of departure for improving the EHR design and redesign.

Blijleven et al [20] developed the Sociotechnical EHR Workaround Analysis (SEWA) framework for identifying, analyzing, and subsequently resolving EHR workarounds. The framework was inspired by the Systems Engineering Initiative for Patient Safety (SEIPS) framework [21]. The SEWA framework incorporates four angles: the different rationales EHR users have for creating workarounds (eg, memory aid and required data entry option missing), the stakeholders affected by a workaround (eg, patient and health care professional), the impact of a workaround (eg, on safety and efficiency), and inherent attributes of workarounds (eg, unavoidable, repetitive, and cascading).

The SEWA framework [20] was based on approximately 200 hours of audiovisual material of user–EHR interaction and

semistructured follow-up interviews in a single large case study in an academic hospital setting [19,22]. However, the authors argued that the applicability of the framework in other contexts might be limited, such as in nonacademic hospitals or in hospitals where paper-based workarounds (eg, for ordering drugs) are still allowed. Therefore, they recommended validation, refinement, and enrichment of the framework by incorporating workarounds and related rationales, attributes, possible scopes, and types of consequences identified in other EHR workaround–related research and clinical contexts.

To address these shortcomings, a scoping literature review was performed to identify and map the available evidence on EHR workarounds [23]. This paper presents a revised version of the SEWA framework, with rationales, attributes, possible scopes, and types of impact described in workaround-related studies in the EHR, electronic medical record, and computerized physician order entry domains in primary, secondary, and tertiary care contexts published between 2010 and 2021.

Methods

Search Strategy

The MEDLINE, Embase, CINAHL, Cochrane, and IEEE databases were searched for relevant studies. We included original, full papers of research with empirical data and conference papers if there were no full papers published in the same study. *Gray literature*, such as books, was not considered. The search queries included the keywords *EHR*, *electronic health record*, and *workaround(s)* and their synonyms. As the aim was to identify new rationales, attributes, consequences, and scopes of EHR workarounds for the enrichment of the SEWA framework, we defined the searches as broad as possible. Pilot literature searches were conducted to check the appropriateness of the queries. During the pilot searches, the term *workflow* was used as a possible synonym for workarounds. The inclusion of this term led to a much larger pool of possible studies. However, most of these studies were focused on care processes that have no relation with EHR use and were thus, out of scope. Therefore, this term was excluded from search queries. Furthermore, to include the complete spectrum of possible EHRs, a combination of the terms *health/medical/patient/health care/clinical record* and *electronic/digital/online* was used. The results of this pilot evaluation were used to adjust the queries. The used queries are shown in [Table 1](#).

Table 1. Search queries used for the scoping review.

Date of search	Database	Query
April 9, 2021	MEDLINE	<i>((([([([([([([health record* OR medical record* OR patient record* OR health care record* OR clinical record* AND electronic] OR digital] OR digitized] OR online) OR online] OR [(Electronic Health Records (MeSH Terms)] OR electronic health record*) OR EHR] OR [(Medical Records Systems, Computerized (MeSH Terms)] OR computerized patient record) OR computerised patient record]) AND ((workaround*) OR work around*)] OR workaround*))</i>
April 9, 2021	Embase	<i>(workaround OR workaround* OR workaround OR workaround*) AND ((health record* OR medical record* OR patient record* OR health care record* OR clinical record*) AND (electronic OR digital OR online OR online OR digitized OR digitised]) OR [electronic health record* OR ehr OR electronic medical record* OR emr] OR [computerized patient record OR computerised patient record])</i>
April 9, 2021	CINAHL	<i>(workaround OR work around OR workarounds) AND ((health record OR medical record OR patient record OR health care record OR clinical record) AND (electronic OR digital OR [online OR online] OR [digitized OR digitised])) OR [electronic health record* OR EHR OR electronic medical record* OR EMR] OR [computerized patient record OR computerised patient record])</i>
April 9, 2021	IEEE	<i>((([([([([([workaround*) OR work around*) OR workaround*])) AND ((health record OR medical record OR patient record OR health care record OR clinical record) AND [electronic OR digital OR (online OR online) OR (digitized OR digitised)]) OR (electronic health record* OR EHR OR electronic medical record* OR EMR) OR (computerized patient record OR computerised patient record)))</i>
April 9, 2021	Cochrane	<i>(workaround*): ti, ab, kw OR (work-around*): ti, ab, kw OR (work around*): ti, ab, kw AND ((electronic health record*): ti, ab, kw OR (health record*): ti, ab, kw OR (medical record*): ti, ab, kw OR (patient record*): ti, ab, kw OR (health care record): ti, ab, kw OR (EHR): ti, ab, kw OR (EMR): ti, ab, kw OR (clinical record): ti, ab, kw OR ([computerized patient record]: ti, ab, kw OR [computerized patient record]: ti, ab, kw]) AND [electronic]: ti, ab, kw OR [digital]: ti, ab, kw OR [online]: ti, ab, kw OR [online]: ti, ab, kw OR [digitized]: ti, ab, kw OR [digitised]: ti, ab, kw)</i>

Selection Criteria

The inclusion and exclusion criteria were chosen through discussions among the reviewers (FH, VB, and MJ). As the focus of this scoping review was on workarounds in EHR use, it was decided to exclude studies focused on barcode medication administration systems as these systems serve only 1 purpose and cover only a small part of the medication process. Furthermore, the choice was made to exclude research focused on EHR functionalities other than those aimed at supporting the clinical process. To ensure data quality, a study was excluded if the research methods were not reported or in case the study had not been peer reviewed. Furthermore, research published before 2010 was excluded as EHRs have undergone significant changes and improvements over the years. Finally, the inclusion and exclusion criteria were chosen.

The study inclusion criteria were as follows:

1. The health care setting of the study must be either ≥ 1 of primary, secondary, or tertiary care.
2. Workarounds were studied or reported in the context of EHR use.

3. The article was published between 2010 and 2021.

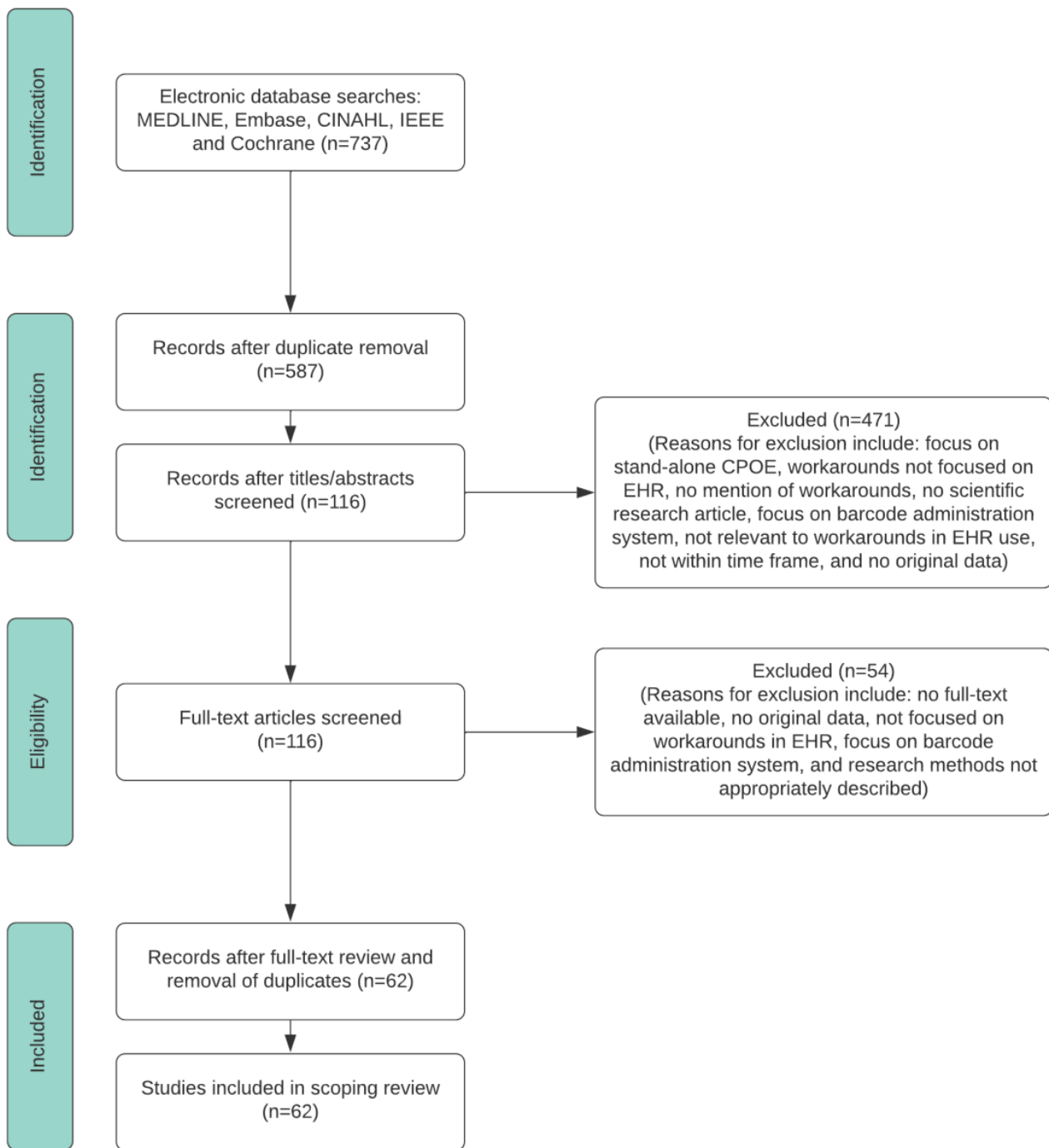
Studies were excluded if they met any of the following criteria:

1. The research focused on EHR functionalities other than those aimed at supporting within the clinical process.
2. The research focused on a barcode administration functionality.
3. The article was not written in English.
4. There was no access to the full-text article.
5. The article was not peer reviewed.
6. The research methods were not reported.

Article Selection

A literature search was conducted in April 2021. A total of 737 potentially relevant studies were retrieved from our initial search of electronic databases, more specifically MEDLINE (263/737, 35.7%), Embase (121/737, 16.4%), CINAHL (89/737, 12.1%), IEEE (58/737, 7.9%), and Cochrane (206/737, 27.9%). The results of the study selection process are shown in the PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flowchart in [Figure 1](#).

Figure 1. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flowchart of the study selection process. CPOE: computerized physician order entry; EHR: electronic health record.



The retrieved 737 studies were uploaded to EndNoteX9 (Clarivate), in which duplicates were first removed by both using EndNoteX9 and by performing a manual check (Figure 1). This led to 79.6% (587/737) of unique studies. These studies were reviewed by two independent reviewers (FH and VB). The 2 reviewers first independently screened the titles and abstracts of the eligible papers to evaluate whether they met the inclusion criteria. Of the 587 studies, 116 (19.8%) studies met the inclusion criteria, and 471 (60.2%) studies were excluded (because of, for example, workarounds not being focused on the EHR, not being a scientific research article, and no workarounds mentioned). Afterward, the reviewers independently screened the full texts of these 116 studies,

leading to 62 (53.4%) included studies and 54 (46.6%) excluded studies (eg, no full-text available and methods inappropriately described). After each screening phase, the two reviewers (FH and VB) discussed their findings. The next screening phase was conducted only if a consensus was reached between the 2 independent reviewers. If a disagreement between the 2 reviewers could not be resolved by discussion, a third independent reviewer (MJ) was involved. After consensus was reached, interrater reliability was reported by calculating the Cohen κ . The interrater agreement was also calculated to show the extent to which the reviewers were able to reconcile through discussion [24]. For the first round (title and abstract screening), the Cohen κ value was 0.958, and the interrater agreement value

was 0.985. For the second round (full-text screening), the Cohen κ value was 0.930, and the interrater agreement value was 0.966.

Data Analysis of Included Articles

Descriptive data from the included articles, such as title, authors, year of publication, study setting, functionalities of EHR studied, and research methods used, were captured in a generic overview per study ([Multimedia Appendix 1](#)). Workaround-related data from the included articles, such as workaround rationales, attributes, consequences, and scope, were captured in an analytic frame per study ([Multimedia Appendix 2](#)).

The data extracted from the included articles were compared with the SEWA framework on a study-by-study basis. In doing so, SEWA was supplemented with new rationales, attributes, possible scopes, and types of impact of EHR workarounds that were not previously included. After the analysis was completed, an updated (graphical) version of the SEWA framework was created.

Results

General Characteristics

The general characteristics of the 62 studies are shown in [Table 2](#). There was an approximately even split in studies published between 2010 and 2015 and between 2016 and 2021. The study settings were almost equally distributed, with most (23/62, 37%) being set in tertiary care, such as academic hospitals and special care units. The largest group of studies (28/62, 45%) focused their research on the EHR overall. Of the 62 studies, 17 (27%) studied medication-related functionalities or EHR-integrated systems, such as computerized physician order entries. Approximately half (28/62, 45%) used or included a combination of physicians, nurses, and other staff such as pharmacists and administrative personnel as participants. Of the 62 studies, 26 (42%) used a combination of methods such as observations, interviews, and questionnaires, 15 (24%) used interviews as the sole method, 5 (8%) solely used questionnaires, 7 (11%) solely used observational methods, and 9 (15%) used other methods such as think-aloud protocols and documentation analysis.

Table 2. General characteristics of the included studies (N=62).

Study characteristics	Values, n (%)
Year of publication	
2010-2015	30 (48)
2016-2021	32 (52)
Study setting	
Primary care	18 (29)
Secondary care	21 (34)
Tertiary care	23 (37)
Functionalities of EHR^a studied	
Medication-related (eg, prescribing and CPOE ^b)	17 (27)
Documentation	8 (13)
Overall EHR	28 (45)
Others (eg, alert systems and authentication process)	9 (15)
Type of population	
Physicians	9 (15)
Nurses	13 (21)
Others (eg, pharmacists or administrative staff such as managers, assistants, secretary, or not mentioned)	12 (19)
Combination of users	28 (45)
Methods	
Observations	7 (11)
Interviews	15 (24)
Questionnaires	5 (8)
Others (eg, think-aloud and documentation analysis)	9 (15)
Combination of ≥ 1 observation, interview, questionnaire, or other	26 (42)

^aEHR: electronic health record.

^bCPOE: computerized physician order entry.

Validation, Refinement, and Enrichment of the SEWA Framework

Overview

Evidence for the work system components, rationales, attributes, type of impact, and possible scopes contained in the original SEWA framework was found in the included studies. Moreover, we refined and enriched the original framework with 7 rationales, 4 attributes, and 3 types of impact. The following subsections elaborate on the work system components, rationales, attributes, possible scopes, and types of impact.

Work System Components

Support for all 5 work system components was found in the included studies, as shown in Table 3. No new work system components were identified. However, we made 1 change to the work system component *EHR system*, which we renamed to *EHR system and related technology*. The latter was incorporated to also cover workarounds stemming from the use of technology other than the EHR but used in parallel with the EHR, such as scanners [25].

Table 3. Overview of work system components and related included studies.

Work components	Description	Studies
Person(s)	Health care professionals developing and using EHR ^a workarounds	[20,26-28]
EHR system and related technology	The EHR and related information technology used by health care professionals	[20,25-27,29-31]
Organization	Organizational conditions (eg, care directives and hospital policies) under which clinical tasks and EHR use are performed	[20,27,28,30,31]
Physical environment	The environment (eg, outpatient examination room and inpatient ward) and its conditions (eg, lighting and noise) in which clinical tasks are conducted by health care professionals	[20,26,27]
Task(s)	Clinical tasks performed by health care professionals	[20,26,28,30-32]

^aEHR: electronic health record.

Rationales

The rationales for workarounds contained in the original SEWA framework were confirmed in many studies. In addition, 7 new rationales were identified.

Under the work system component *person(s)*, one rationale was added: *trust* (Table 4). Multiple studies reported that users created workarounds because of insufficient trust in the (new) system or its capabilities while frequently maintaining trust in older systems (replaced by the EHR). The related causes of a lack of trust are a lack of perceived usefulness of the (new) system and insufficient confidence in (completeness) of the data available in the EHR [33-39]. The description of the rationale *awareness* has been refined to also cover awareness of the information needs of patients and not just of colleagues [40]. Likewise, the description of the rationale *social norms* has been refined to make cultural [30,41] and collaborative [27,42] aspects more explicit.

Although extensive support in the included studies was found for all rationales under the work system component *EHR system and related technology*, except *patient data specificity*, four additional rationales were identified: *data integration*, *enforced actions*, *data quality*, and *interoperability* (Table 5). The

description of the pre-existing rationale *technical issues* has been refined to cover technical issues related to ancillary technology used in conjunction with the EHR.

Multiple studies provide support for all rationales under the work system component *organization* except for the rationale *data migration policy* (Table 6). No new rationales were identified.

Although support was found for the pre-existing rationales under *task(s)*, one rationale was added: *task complexity* (Table 7). Approximately 3% (2/62) of studies described that the EHR does not always sufficiently support the execution of a complex task at hand [34-39]. Therefore, health care professionals resort to workarounds to make their workflow more digestible.

Finally, the SEIPS work system component *physical environment* was incorporated into the original SEWA framework without any rationale. However, Dudding et al [25] mentioned that a busy, fast-paced environment where interruptions are constant, such as the neonatal intensive care unit, gives rise to EHR workarounds. The rationale here is “fast-paced environment” and is described as “devising workarounds to cope with the inability to, for example, update the documentation in fast-paced care environments where interruptions are constant” [25].

Table 4. Overview of rationales for the work system component person(s) and related included studies.

Rationales	Description	Studies
Declarative knowledge	Not knowing how to use (a part of) the EHR ^a to accomplish a task	[20,33,34,39,43,44]
Procedural knowledge	Knowing how but not being proficient enough to use a part of the EHR to accomplish a task	[20,28,34,39,44]
Memory aid	Writing patient data down on paper (eg, keywords) or adding visual elements to parts of text in a progress note (eg, boldfacing, italicizing, or underlining) to remind oneself	[20,34,39,43,45-47]
Awareness	Storing patient data that are perceived important by the EHR user for other colleagues or patients to be noticed (frequently in a data field other than the intended field in the EHR)	[20,39,40,48]
Social norms	Formal or informal, collaborative, and cultural understandings among health care professionals leading to the creation and dissemination of workarounds (eg, mimicking workarounds devised by colleagues to accomplish a task or working around the system upon as friendly requested or enforced by a fellow clinician)	[20,29-31,45,49,50]
Trust (new)	Having insufficient trust in the (new) EHR system or its capabilities, lack of perceived usefulness, or insufficient confidence in the (completeness) of data	[20,33-39]

^aEHR: electronic health record.

Table 5. Overview of rationales for the work system component EHR^a system and related technology and related included studies.

Rationales	Description	Studies
Usability	High behavioral user cost in accomplishing a task	[20,25,28,29,31,41,42,45,46,50-56]
Technical issues	(A part of the) EHR or ancillary technology halting, crashing, or slowing down, hampering the EHR user in accomplishing a task	[20,25,28,31-33,43,44,51-53,55-61]
Data presentation	Preferring a different data view (eg, visualization by means of charts or graphs rather than plain text)	[20,55,62]
Patient data specificity	Needing to enter or request patient data with greater or lesser specificity than offered or enforced by the EHR	[20]
Data integration (new)	EHR not providing or supporting the integration of patient data necessary for care delivery	[42,45]
Enforced actions (new)	Avoiding or overriding actions enforced by the EHR (eg, bypassing the approval process of prescribing medication or using a different user account)	[29,43,48,54,63]
Data quality (new)	Unavailability of data, disparity in data formats (eg, the same data being stored in multiple different formats in the EHR), lack of standardization, and information gaps in the EHR	[31,34-36,39,41,42,44,50,57,64-67]
Interoperability (new)	Data not able to be exchanged between health care systems or institutions (eg, causing data to be unavailable at the right moment and time)	[44,50,54,56,64,65]

^aEHR: electronic health record.

Table 6. Overview of rationales for the work system component organization and related included studies.

Rationales	Description	Studies
Efficiency	Using an alternative way of accomplishing a task that improves actual efficiency	[20,29,31,34,35,37,43,46,47,55,68-70]
Data migration policy	Not having (direct) access to required historical data because of data not having been imported from previously used systems to the current EHR ^a	[20]
Enforced data entry	EHR enforcing user to enter patient data of which neither the user nor the patient has knowledge of	[20,71,72]
Required data entry option missing	EHR not offering the required data entry option (eg, 3.75 mg rather than the available options 2.5 mg or 5 mg)	[20,32,71]

^aEHR: electronic health record.

Table 7. Overview of rationales for the work system component task(s) and related included studies.

Rationales	Description	Studies
Task interference	Inability to perform multiple tasks at once (eg, simultaneously treating a patient on the treatment table as well as entering patient data into the EHR ^a)	[20,61]
Commitment to patient interaction	Valuing patient interaction over computer interaction (ie, writing things down on paper and afterward entering this into the EHR)	[20,34,37,41,44,55,61,73]
Task complexity (new)	The high complexity of the tasks needing to be conducted	[34,39]

^aEHR: electronic health record.

Attributes

Although several studies confirmed the previously defined attributes in SEWA, several included studies also mentioned a total of 4 new attributes (Table 8). These are concerned with whether the user is aware of using a workaround [49]

(*awareness*), whether the workaround is an individual or shared practice across users [49] (*shared*), on what medium the workaround is conducted (eg, paper or computer) [34,41] (*medium*), and whether the workaround is a formal or informal practice (eg, part of a defined process or approved or promoted by management or not) [56] (*formality*).

Table 8. Overview of workaround attributes and related included studies.

Attributes	Description	Source
Cascadedness	Whether the workaround initiates the creation of 1 or multiple additional workarounds or is an isolated occurrence	[20]
Avoidability	Whether the workaround is required to proceed with one’s workflow or optional	[20,32,66,74]
Anticipatedness	Whether the workaround is used at known moments in time (ie, the situation in which the workaround is used is known beforehand) or used unexpectedly	[20,74]
Repetitiveness	Whether the workaround is ingrained into the workflow (ie, becomes part of daily routines) or used temporarily to overcome workflow constraints	[20,56,74]
Awareness (new)	Whether the user is aware of using the workaround	[49]
Shared (new)	Whether the workaround is a shared practice across multiple other users of the EHR ^a or limited to 1 user	[49]
Medium (new)	On what medium the workaround is conducted (eg, paper, computer, verbal, or a combination)	[34,41]
Formality (new)	Whether the use of the workaround is approved by management and part of a defined process	[56]

^aEHR: electronic health record.

Types of Impact

The previously defined types of impact in the SEWA framework were confirmed by many included studies. Multiple additional types of impact were also identified: *privacy/security*, *data quality*, *employee perception of EHR*, *financial*, *law/regulations*, and *workload* (Table 9). *Privacy/security* relates not only to the impact a workaround has on the security and privacy of the data

but also to the patient and organization itself. Data quality concerns the impact on, for example, loss of data, or a lower data quality because of spelling or formatting mistakes in the data. Moreover, workarounds can have a positive or negative financial impact [58], may jeopardize laws and regulations [63,75], and have a positive or negative impact on the workload of the user [43].

Table 9. Overview of types of impact and related included studies.

Impact	Description	Source
Patient safety	The impact on the safety (physical and mental) of the patient	[20,28,29,41,43,46,48,53,54,58,59,67,75-77]
Effectiveness of care	The effectiveness and quality of the care process performed	[20,28,43,46,54,58,59,67]
Efficiency of care	The impact on the efficiency of the care process in terms of time and resources expended	[20,33,55,60,64,72,76]
Privacy and security (new)	Impact on the security and privacy of data related to the patient or organization	[32,39,51,52,56,63,68,75]
Data quality (new)	Impact of workarounds on data quality (eg, loss of data or decreased data quality)	[32,33,35,39,41,46,51,52,56,59,69,76]
Financial (new)	Financial implications because of the workaround	[58]
Laws and regulations (new)	Legal conflicts resulting from the use of a workaround	[63,75]
Workload (new)	An increase or decrease in workload of the EHR ^a user resulting from the use of a workaround	[43]

^aEHR: electronic health record.

Possible Scopes

Only a few studies explicitly discussed possible scopes (ie, entities impacted) of workarounds and resonated with those in

the SEWA framework [41,43,53,77] (Table 10). No new possible scopes were identified.

Table 10. Overview of possible scopes and related included studies.

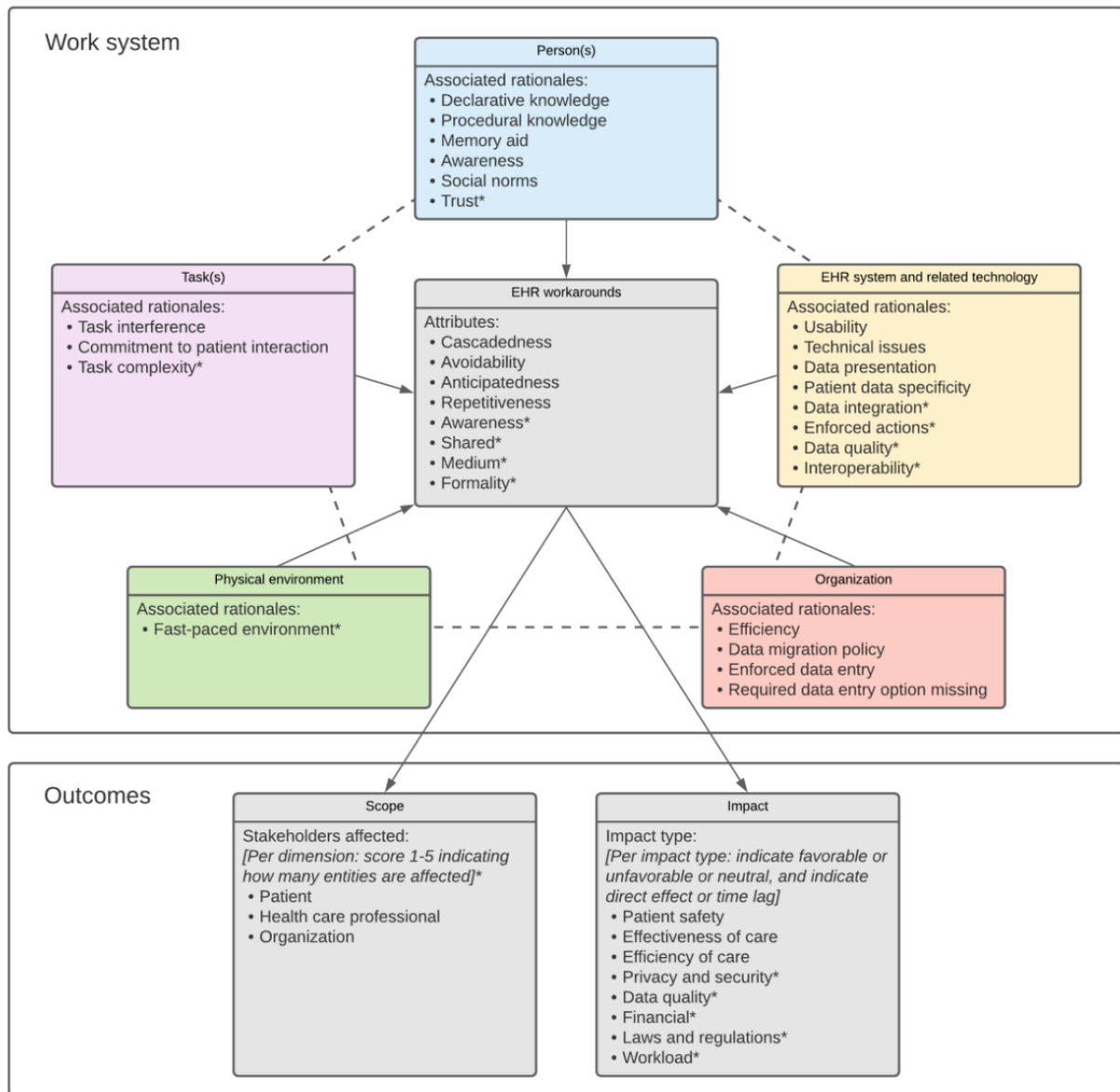
Scope	Description	Source
Patient	The workaround affects the patients in the care process	[20,43,77]
Health care professional	The workaround affects the health care professionals such as physicians, nurses, and pharmacists	[20]
Organization	The workaround affects the whole organization, including the supporting departments such as finance or legal	[20,41,53]

Revised Version of the SEWA Framework

On the basis on the foregoing results, the original SEWA framework [20] was revised to incorporate new rationales, attributes, types of impact, and possible scopes identified in the included studies (Figure 2). The revised SEWA framework still comprises 2 major parts. The first part concerns the work system and its components (inspired by the SEIPS framework), [21] constituting the context in which EHR workarounds are created.

The work system components now include 22 rationales (previously 15) for workaround creation, and EHR workarounds are now defined by 8 attributes (previously 4). The second part concerns the possible scope of workarounds in terms of types and number of entities affected (still 3), as well as their impact on patient safety, the effectiveness of care, the efficiency of care, and 5 newly introduced types of impact. All new items in the framework are marked with asterisks.

Figure 2. Revised SEWA framework with incorporated rationales, attributes, types of impact, and possible scopes identified in included studies. EHR: electronic health record; SEWA: Sociotechnical Electronic Health Record Workaround Analysis.



The recommendations [20] for using a scoring mechanism to indicate whether the impact per workaround is favorable, unfavorable, or neutral, as well as to indicate whether the impact is immediate or only observable after a certain period (*direct/time lag*) remain. However, we also recommend the inclusion of a scoring mechanism to indicate the number of patients and health care professionals and organizational units affected per applicable scope. This is in line with Carayon et al [53], who distinguished between workarounds having an impact at an individual or *team level* (eg, an entire team of nurses in a

certain hospital ward). Applying a scoring mechanism allows for a more substantiated view when analyzing and prioritizing various identified workarounds for resolution.

Discussion

Principal Findings

A scoping review was performed to theoretically validate the SEWA framework [20] and refine and enrich it with newly identified rationales, attributes, types of impact, and possible

scopes of EHR workarounds. The scoping review retrieved 737 studies, of which 62 (8.4%) were included. The included studies provided extensive support for nearly all the items included in the original SEWA framework. SEWA was revised and enriched with 7 new rationales, 4 attributes, and 5 types of impact of EHR workarounds mentioned in the included studies. The definitions of several existing rationales were also refined. As a result, SEWA is now grounded in the existing body of peer-reviewed empirical evidence on EHR workarounds published between 2010 and 2021. In addition, this revised version is likely also applicable in a wider range of health care settings as input for the original SEWA framework that came from a single comprehensive case study on EHR workarounds in an academic hospital.

Comparison With the Literature

The results of this scoping review are in line with prior research and reviews of EHR workarounds. In an integrative review, Fraczkowski et al [78] examined nurse workarounds in EHR use. The categories defined in the review by Fraczkowski et al [78] are similar to the work system components defined in SEWA, with the exception of *usability* being a separate rationale in the SEWA framework under the work system component *EHR system and related technology* [20]. The *patient* category in the review by Fraczkowski et al [78] is defined as an impact and scope category in SEWA [20]. Finally, Fraczkowski et al [78], similar to Koppel et al [18], did not include a work system component for *person(s)* (the users of the EHR) as a category. Our scoping review is one of the few studies that investigated the entire spectrum of EHR users. On the one hand, we included studies of all types of health care professionals in primary, secondary, and tertiary care who make use of an EHR in their clinical practice, whereas other reviews merely focused on a specific population such as physicians, nurses, or secretary personnel [78]. On the other hand, we excluded studies researching workarounds in the use of barcode medication administration systems, whereas other reviews did not [78].

Strengths and Limitations

To maximize the capture of relevant information on EHR workarounds, comprehensive and structured searches were conducted in MEDLINE, Embase, CINAHL, Cochrane, and IEEE databases. Data charting templates and analytic frames were used to extract relevant information from the reviewed studies and compare with pre-existing items in the SEWA framework.

A total of 2 research team members participated in the review process for both the title and abstract and full-text review phases, with a Cohen κ value of >0.9 . This indicates an adequate interrater agreement. Despite this, our scoping review is at risk for selection bias, as we did not identify all available data, such as gray literature on EHR workarounds. There is a chance that relevant but nonincluded studies may use terminology other than the terms included in the search queries.

The broad scope of the retrieved information on EHR workarounds and the different types of studies reporting a particular issue made using a formal meta-analytic method to quantitatively assess the quality of the studies and evidence of

retrieved information difficult. However, given the purpose of the scoping review to theoretically validate and refine the SEWA framework, we do not consider this limitation.

Implications for Practice and Future Research

Multidisciplinary teams (comprising, for example, physicians, nurses, management, and EHR developers) can use the revised SEWA framework to identify, analyze, prioritize, and resolve workarounds related to EHR use more accurately. Similarly, the consequences of current and future configurations of the work system (health care professionals' work processes and activities in relation to their EHR use) can be assessed and discussed in greater detail to determine how a design and redesign of the work system would positively or negatively affect the interaction between work system components. Finally, as workarounds are subject to gradual change (eg, personal changes in experience with the EHR, system updates to the EHR, and hospital policies), more detailed snapshots of the work system using SEWA can be taken over time and compared so as to gain valuable insights into how EHR workarounds evolve over time.

Concerning future research, EHR systems are continuously subject to technological evolution by developments in, for example, artificial intelligence, machine learning, and telemedicine. This may lead to the creation of hitherto unidentified rationales, attributes, possible scopes, and types of impact of workarounds on users, patients, and health care organizations. Similarly, more studies on EHR workarounds will continue to emerge that may report novel insights not incorporated into the revised SEWA framework. Therefore, we expect that SEWA needs a continuous process of refinement over time. This could be done by repeating the scoping review using the described search strategy, search queries, and inclusion and exclusion criteria.

In addition, although the revised SEWA framework is now theoretically validated, refined, and enriched, practical validation is still required. The same holds true when investigating its practicality. The firsthand experience from the application of SEWA in practice could yield suggestions for further improvement. A related suggestion is that although the framework helps in identifying and analyzing workarounds, a prioritization method for handling these issues is likely required, as workarounds are generally abundant in any organization, and resources to resolve them are finite. Therefore, the framework could benefit from being extended with prioritization mechanisms and weighting factors for deciding which workarounds require priority. Similarly, the framework could be translated into a practical tool such as a scoring matrix to facilitate use by practitioners.

Finally, the applicability of the SEWA framework could be explored for systems other than EHRs (eg, enterprise resource planning, customer relationship management, and content management) and in other settings (eg, nonacademic hospitals and general practitioner practices) and even in other industries (eg, financial services and manufacturing) after appropriate validation. Although SEWA has an explicit focus on EHRs used in health care, we expect many of the described workarounds

rationales and attributes to be applicable to other systems, settings, and industries.

Authors' Contributions

VB, FH, and MJ conceived and designed the study. FH collected the data. VB and FH analyzed the data and wrote the manuscript. VB and MJ edited the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Descriptive data template that was captured per included study.

[\[DOCX File, 13 KB - jmir_v24i3e33046_app1.docx\]](#)

Multimedia Appendix 2

Analytic frame with workaround-related data captured per study.

[\[DOCX File, 13 KB - jmir_v24i3e33046_app2.docx\]](#)

References

1. Howe JL, Adams KT, Hettinger AZ, Ratwani RM. Electronic health record usability issues and potential contribution to patient harm. *J Am Med Assoc* 2018 Mar 27;319(12):1276-1278 [[FREE Full text](#)] [doi: [10.1001/jama.2018.1171](https://doi.org/10.1001/jama.2018.1171)] [Medline: [29584833](https://pubmed.ncbi.nlm.nih.gov/29584833/)]
2. 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 Aff (Millwood)* 2018 Nov;37(11):1752-1759. [doi: [10.1377/hlthaff.2018.0699](https://doi.org/10.1377/hlthaff.2018.0699)] [Medline: [30395517](https://pubmed.ncbi.nlm.nih.gov/30395517/)]
3. Melnick ER, Dyrbye LN, Sinsky CA, Trockel M, West CP, Nedelec L, et al. The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clin Proc* 2020 Mar;95(3):476-487 [[FREE Full text](#)] [doi: [10.1016/j.mayocp.2019.09.024](https://doi.org/10.1016/j.mayocp.2019.09.024)] [Medline: [31735343](https://pubmed.ncbi.nlm.nih.gov/31735343/)]
4. Shanafelt TD, Dyrbye LN, West CP. Addressing physician burnout: the way forward. *J Am Med Assoc* 2017 Mar 07;317(9):901-902. [doi: [10.1001/jama.2017.0076](https://doi.org/10.1001/jama.2017.0076)] [Medline: [28196201](https://pubmed.ncbi.nlm.nih.gov/28196201/)]
5. Akbar F, Mark G, Prausnitz S, Warton EM, East JA, Moeller MF, et al. Physician stress during electronic health record inbox work: in situ measurement with wearable sensors. *JMIR Med Inform* 2021 Apr 28;9(4):e24014 [[FREE Full text](#)] [doi: [10.2196/24014](https://doi.org/10.2196/24014)] [Medline: [33908888](https://pubmed.ncbi.nlm.nih.gov/33908888/)]
6. Tai-Seale M, Olson CW, Li J, Chan AS, Morikawa C, Durbin M, et al. Electronic health record logs indicate that physicians split time evenly between seeing patients and desktop medicine. *Health Aff (Millwood)* 2017 Apr 01;36(4):655-662 [[FREE Full text](#)] [doi: [10.1377/hlthaff.2016.0811](https://doi.org/10.1377/hlthaff.2016.0811)] [Medline: [28373331](https://pubmed.ncbi.nlm.nih.gov/28373331/)]
7. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med* 2016 Sep 06;165(11):753. [doi: [10.7326/m16-0961](https://doi.org/10.7326/m16-0961)]
8. Tsou A, Lehmann C, Michel J, Solomon R, Possanza L, Gandhi T. Safe practices for copy and paste in the EHR: systematic review, recommendations, and novel model for health IT collaboration. *Appl Clin Inform* 2017 Dec 20;26(01):12-34. [doi: [10.4338/aci-2016-09-r-0150](https://doi.org/10.4338/aci-2016-09-r-0150)]
9. Sittig DF, Wright A, Ash J, Singh H. New unintended adverse consequences of electronic health records. *Yearb Med Inform* 2018 Mar 06;25(01):7-12. [doi: [10.15265/iy-2016-023](https://doi.org/10.15265/iy-2016-023)]
10. Campbell EM, Sittig DF, Ash JS, Guappone KP, Dykstra RH. Types of unintended consequences related to computerized provider order entry. *J Am Med Inform Assoc* 2006 Sep 01;13(5):547-556. [doi: [10.1197/jamia.m2042](https://doi.org/10.1197/jamia.m2042)]
11. Ash JS, Sittig DF, Dykstra RH, Guappone K, Carpenter JD, Seshadri V. Categorizing the unintended sociotechnical consequences of computerized provider order entry. *Int J Med Inform* 2007 Jun;76 Suppl 1:21-27. [doi: [10.1016/j.ijmedinf.2006.05.017](https://doi.org/10.1016/j.ijmedinf.2006.05.017)] [Medline: [16793330](https://pubmed.ncbi.nlm.nih.gov/16793330/)]
12. Ash JS, Sittig DF, Poon EG, Guappone K, Campbell E, Dykstra RH. The extent and importance of unintended consequences related to computerized provider order entry. *J Am Med Inform Assoc* 2007 Jul 01;14(4):415-423. [doi: [10.1197/jamia.m2373](https://doi.org/10.1197/jamia.m2373)]
13. Campbell EM, Guappone KP, Sittig DF, Dykstra RH, Ash JS. Computerized provider order entry adoption: implications for clinical workflow. *J Gen Intern Med* 2009 Jan 20;24(1):21-26 [[FREE Full text](#)] [doi: [10.1007/s11606-008-0857-9](https://doi.org/10.1007/s11606-008-0857-9)] [Medline: [19020942](https://pubmed.ncbi.nlm.nih.gov/19020942/)]
14. Greenhalgh T, Potts H, Wong G, Bark P, Swinglehurst D. Tensions and paradoxes in electronic patient record research: a systematic literature review using the meta-narrative method. *Milbank Q* 2009 Dec;87(4):729-788 [[FREE Full text](#)] [doi: [10.1111/j.1468-0009.2009.00578.x](https://doi.org/10.1111/j.1468-0009.2009.00578.x)] [Medline: [20021585](https://pubmed.ncbi.nlm.nih.gov/20021585/)]

15. Blijleven V, Koelemeijer K, Jaspers M. Identifying and eliminating inefficiencies in information system usage: a lean perspective. *Int J Med Inform* 2017 Nov;107:40-47. [doi: [10.1016/j.ijmedinf.2017.08.005](https://doi.org/10.1016/j.ijmedinf.2017.08.005)] [Medline: [29029690](https://pubmed.ncbi.nlm.nih.gov/29029690/)]
16. Halbesleben J, Wakefield D, Wakefield B. Work-arounds in health care settings: literature review and research agenda. *Health Care Manage Rev* 2008;33(1):2-12. [doi: [10.1097/01.HMR.0000304495.95522.ca](https://doi.org/10.1097/01.HMR.0000304495.95522.ca)] [Medline: [18091439](https://pubmed.ncbi.nlm.nih.gov/18091439/)]
17. Kobayashi M, Fussell S, Xiao Y, Seagull F. Work coordination, workflow, and workarounds in a medical context. In: *Proceedings of the CHI '05 Extended Abstracts on Human Factors in Computing Systems*. 2005 Presented at: CHI05: CHI 2005 Conference on Human Factors in Computing Systems; April 2 - 7, 2005; Portland OR USA p. 1561-1564. [doi: [10.1145/1056808.1056966](https://doi.org/10.1145/1056808.1056966)]
18. Koppel R, Wetterneck T, Telles JL, Karsh B. Workarounds to barcode medication administration systems: their occurrences, causes, and threats to patient safety. *J Am Med Inform Assoc* 2008 Jul 01;15(4):408-423. [doi: [10.1197/jamia.m2616](https://doi.org/10.1197/jamia.m2616)]
19. Blijleven V, Koelemeijer K, Wetzels M, Jaspers M. Workarounds emerging from electronic health record system usage: consequences for patient safety, effectiveness of care, and efficiency of care. *JMIR Hum Factors* 2017 Oct 05;4(4):e27 [FREE Full text] [doi: [10.2196/humanfactors.7978](https://doi.org/10.2196/humanfactors.7978)] [Medline: [28982645](https://pubmed.ncbi.nlm.nih.gov/28982645/)]
20. Blijleven V, Koelemeijer K, Jaspers M. SEWA: a framework for sociotechnical analysis of electronic health record system workarounds. *Int J Med Inform* 2019 May;125:71-78. [doi: [10.1016/j.ijmedinf.2019.02.012](https://doi.org/10.1016/j.ijmedinf.2019.02.012)] [Medline: [30914183](https://pubmed.ncbi.nlm.nih.gov/30914183/)]
21. 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 Nov;56(11):1669-1686 [FREE Full text] [doi: [10.1080/00140139.2013.838643](https://doi.org/10.1080/00140139.2013.838643)] [Medline: [24088063](https://pubmed.ncbi.nlm.nih.gov/24088063/)]
22. Blijleven V, Koelemeijer K, Jaspers M. Exploring workarounds related to electronic health record system usage: a study protocol. *JMIR Res Protoc* 2017 Apr 28;6(4):e72 [FREE Full text] [doi: [10.2196/resprot.6766](https://doi.org/10.2196/resprot.6766)] [Medline: [28455273](https://pubmed.ncbi.nlm.nih.gov/28455273/)]
23. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol* 2018 Nov 19;18(1):143 [FREE Full text] [doi: [10.1186/s12874-018-0611-x](https://doi.org/10.1186/s12874-018-0611-x)] [Medline: [30453902](https://pubmed.ncbi.nlm.nih.gov/30453902/)]
24. Campbell JL, Quincy C, Osserman J, Pedersen OK. Coding in-depth semistructured interviews: problems of unitization and intercoder reliability and agreement. *Sociolog Methods Res* 2013 Aug 21;42(3):294-320. [doi: [10.1177/0049124113500475](https://doi.org/10.1177/0049124113500475)]
25. Dudding K, Gephart S, Carrington J. Neonatal nurses experience unintended consequences and risks to patient safety with electronic health records. *Comput Inform Nurs* 2018 Apr;36(4):167-176. [doi: [10.1097/CIN.0000000000000406](https://doi.org/10.1097/CIN.0000000000000406)] [Medline: [29194056](https://pubmed.ncbi.nlm.nih.gov/29194056/)]
26. Niazkhani Z, Pirnejad H, van der Sijs H, Aarts J. Evaluating the medication process in the context of CPOE use: the significance of working around the system. *Int J Med Inform* 2011 Jul;80(7):490-506. [doi: [10.1016/j.ijmedinf.2011.03.009](https://doi.org/10.1016/j.ijmedinf.2011.03.009)] [Medline: [21555237](https://pubmed.ncbi.nlm.nih.gov/21555237/)]
27. Holden RJ. What stands in the way of technology-mediated patient safety improvements?: a study of facilitators and barriers to physicians' use of electronic health records. *J Patient Saf* 2011 Dec;7(4):193-203 [FREE Full text] [doi: [10.1097/PTS.0b013e3182388cfa](https://doi.org/10.1097/PTS.0b013e3182388cfa)] [Medline: [22064624](https://pubmed.ncbi.nlm.nih.gov/22064624/)]
28. Salahuddin L, Ismail Z, Abd Ghani MK, Aboobaidier BM, Hasan Basari AS. Exploring the contributing factors to workarounds to the hospital information system in Malaysian hospitals. *J Eval Clin Pract* 2020 Oct 20;26(5):1416-1424. [doi: [10.1111/jep.13326](https://doi.org/10.1111/jep.13326)] [Medline: [31863517](https://pubmed.ncbi.nlm.nih.gov/31863517/)]
29. Baysari M, Gigante JD, Moran M, Sandaradura I, Li L, Richardson K, et al. Redesign of computerized decision support to improve antimicrobial prescribing. *Appl Clin Inform* 2017 Dec 20;08(03):949-963. [doi: [10.4338/aci2017040069](https://doi.org/10.4338/aci2017040069)]
30. Ser G, Robertson A, Sheikh A. A qualitative exploration of workarounds related to the implementation of national electronic health records in early adopter mental health hospitals. *PLoS One* 2014;9(1):e77669 [FREE Full text] [doi: [10.1371/journal.pone.0077669](https://doi.org/10.1371/journal.pone.0077669)] [Medline: [24454678](https://pubmed.ncbi.nlm.nih.gov/24454678/)]
31. Stevenson JE, Israelsson J, Petersson G, Bath PA. Factors influencing the quality of vital sign data in electronic health records: a qualitative study. *J Clin Nurs* 2018 Mar 08;27(5-6):1276-1286. [doi: [10.1111/jocn.14174](https://doi.org/10.1111/jocn.14174)] [Medline: [29149483](https://pubmed.ncbi.nlm.nih.gov/29149483/)]
32. Ali M, Cornford T, Klecun E. Exploring control in health information systems implementation. *Stud Health Technol Inform* 2010;160(Pt 1):681-685. [Medline: [20841773](https://pubmed.ncbi.nlm.nih.gov/20841773/)]
33. Crosson JC, Etz RS, Wu S, Straus SG, Eisenman D, Bell DS. Meaningful use of electronic prescribing in 5 exemplar primary care practices. *Ann Fam Med* 2011 Sep 12;9(5):392-397 [FREE Full text] [doi: [10.1370/afm.1261](https://doi.org/10.1370/afm.1261)] [Medline: [21911757](https://pubmed.ncbi.nlm.nih.gov/21911757/)]
34. Flanagan ME, Saleem JJ, Millitello LG, Russ AL, Doebbeling BN. Paper- and computer-based workarounds to electronic health record use at three benchmark institutions. *J Am Med Inform Assoc* 2013 Jun 01;20(e1):59-66 [FREE Full text] [doi: [10.1136/amiajnl-2012-000982](https://doi.org/10.1136/amiajnl-2012-000982)] [Medline: [23492593](https://pubmed.ncbi.nlm.nih.gov/23492593/)]
35. Horsky J, Drucker E, Ramelson H. Higher accuracy of complex medication reconciliation through improved design of electronic tools. *J Am Med Inform Assoc* 2018 May 01;25(5):465-475 [FREE Full text] [doi: [10.1093/jamia/ocx127](https://doi.org/10.1093/jamia/ocx127)] [Medline: [29121197](https://pubmed.ncbi.nlm.nih.gov/29121197/)]
36. Djalali S, Ursprung N, Rosemann T, Senn O, Tandjung R. Undirected health IT implementation in ambulatory care favors paper-based workarounds and limits health data exchange. *Int J Med Inform* 2015 Nov;84(11):920-932. [doi: [10.1016/j.ijmedinf.2015.08.001](https://doi.org/10.1016/j.ijmedinf.2015.08.001)] [Medline: [26296433](https://pubmed.ncbi.nlm.nih.gov/26296433/)]

37. Watson A, Skipper C, Steury R, Walsh H, Levin A. Inpatient nursing care and early warning scores: a workflow mismatch. *J Nurs Care Qual* 2014;29(3):215-222. [doi: [10.1097/NCQ.0000000000000058](https://doi.org/10.1097/NCQ.0000000000000058)] [Medline: [24569518](https://pubmed.ncbi.nlm.nih.gov/24569518/)]
38. Hysong SJ, Sawhney MK, Wilson L, Sittig DF, Espadas D, Davis T, et al. Provider management strategies of abnormal test result alerts: a cognitive task analysis. *J Am Med Inform Assoc* 2010 Jan 01;17(1):71-77. [doi: [10.1197/jamia.m3200](https://doi.org/10.1197/jamia.m3200)]
39. Saleem JJ, Russ AL, Neddo A, Blades PT, Doebbeling BN, Foresman BH. Paper persistence, workarounds, and communication breakdowns in computerized consultation management. *Int J Med Inform* 2011 Jul;80(7):466-479. [doi: [10.1016/j.ijmedinf.2011.03.016](https://doi.org/10.1016/j.ijmedinf.2011.03.016)] [Medline: [21530383](https://pubmed.ncbi.nlm.nih.gov/21530383/)]
40. Ihlebæk HM. Lost in translation - Silent reporting and electronic patient records in nursing handovers: an ethnographic study. *Int J Nurs Stud* 2020 Sep;109:103636 [FREE Full text] [doi: [10.1016/j.ijnurstu.2020.103636](https://doi.org/10.1016/j.ijnurstu.2020.103636)] [Medline: [32593879](https://pubmed.ncbi.nlm.nih.gov/32593879/)]
41. Stevenson JE, Israelsson J, Nilsson G, Petersson G, Bath PA. Vital sign documentation in electronic records: the development of workarounds. *Health Informatics J* 2016 Aug 19;206-215. [doi: [10.1177/1460458216663024](https://doi.org/10.1177/1460458216663024)] [Medline: [27542887](https://pubmed.ncbi.nlm.nih.gov/27542887/)]
42. Carrington J, Effken J. Strengths and limitations of the electronic health record for documenting clinical events. *Comput Inform Nurs* 2011 Jun;29(6):360-367. [doi: [10.1097/NCN.0b013e3181fc4139](https://doi.org/10.1097/NCN.0b013e3181fc4139)] [Medline: [21107239](https://pubmed.ncbi.nlm.nih.gov/21107239/)]
43. Lee S, Lee M. Nurses' electronic medical record workarounds in a tertiary teaching hospital. *Comput Inform Nurs* 2021 Mar 08;39(7):367-374. [doi: [10.1097/CIN.0000000000000692](https://doi.org/10.1097/CIN.0000000000000692)] [Medline: [33675300](https://pubmed.ncbi.nlm.nih.gov/33675300/)]
44. Rathert C, Porter TH, Mittler JN, Fleig-Palmer M. Seven years after meaningful use: physicians' and nurses' experiences with electronic health records. *Health Care Manage Rev* 2019;44(1):30-40. [doi: [10.1097/hmr.0000000000000168](https://doi.org/10.1097/hmr.0000000000000168)]
45. Assis-Hassid S, Grosz BJ, Zimlichman E, Rozenblum R, Bates DW. Assessing EHR use during hospital morning rounds: a multi-faceted study. *PLoS One* 2019 Feb 25;14(2):e0212816 [FREE Full text] [doi: [10.1371/journal.pone.0212816](https://doi.org/10.1371/journal.pone.0212816)] [Medline: [30802267](https://pubmed.ncbi.nlm.nih.gov/30802267/)]
46. Cresswell KM, Worth A, Sheikh A. Integration of a nationally procured electronic health record system into user work practices. *BMC Med Inform Decis Mak* 2012 Mar 08;12(1):15 [FREE Full text] [doi: [10.1186/1472-6947-12-15](https://doi.org/10.1186/1472-6947-12-15)] [Medline: [22400978](https://pubmed.ncbi.nlm.nih.gov/22400978/)]
47. Menon S, Murphy D, Singh H, Meyer AN, Sittig D. Workarounds and test results follow-up in electronic health record-based primary care. *Appl Clin Inform* 2017 Dec 16;07(02):543-559. [doi: [10.4338/aci-2015-10-ra-0135](https://doi.org/10.4338/aci-2015-10-ra-0135)]
48. Van Der Sijs H, Rootjes I, Aarts J. The shift in workarounds upon implementation of computerized physician order entry. *Stud Health Technol Inform* 2011;169:290-294. [Medline: [21893759](https://pubmed.ncbi.nlm.nih.gov/21893759/)]
49. Barrett AK, Stephens KK. Making Electronic Health Records (EHRs) work: informal talk and workarounds in healthcare organizations. *Health Commun* 2017 Aug 27;32(8):1004-1013. [doi: [10.1080/10410236.2016.1196422](https://doi.org/10.1080/10410236.2016.1196422)] [Medline: [27463257](https://pubmed.ncbi.nlm.nih.gov/27463257/)]
50. O'Malley AS, Grossman JM, Cohen GR, Kemper NM, Pham HH. Are electronic medical records helpful for care coordination? Experiences of physician practices. *J Gen Intern Med* 2010 Mar 22;25(3):177-185 [FREE Full text] [doi: [10.1007/s11606-009-1195-2](https://doi.org/10.1007/s11606-009-1195-2)] [Medline: [20033621](https://pubmed.ncbi.nlm.nih.gov/20033621/)]
51. Bowen JL, O'Brien BC, Ilgen JS, Irby DM, Cate OT. Chart stalking, list making, and physicians' efforts to track patients' outcomes after transitioning responsibility. *Med Educ* 2018 Apr 30;52(4):404-413. [doi: [10.1111/medu.13509](https://doi.org/10.1111/medu.13509)] [Medline: [29383741](https://pubmed.ncbi.nlm.nih.gov/29383741/)]
52. Bristol A, Nibbelink C, Gephart S, Carrington J. Nurses' use of positive deviance when encountering electronic health records-related unintended consequences. *Nurs Adm Q* 2018;42(1):1-11. [doi: [10.1097/NAQ.0000000000000264](https://doi.org/10.1097/NAQ.0000000000000264)] [Medline: [29194338](https://pubmed.ncbi.nlm.nih.gov/29194338/)]
53. Carayon P, Hundt AS, Hoonakker P. Technology barriers and strategies in coordinating care for chronically ill patients. *Appl Ergon* 2019 Jul;78:240-247 [FREE Full text] [doi: [10.1016/j.apergo.2019.03.009](https://doi.org/10.1016/j.apergo.2019.03.009)] [Medline: [31046955](https://pubmed.ncbi.nlm.nih.gov/31046955/)]
54. Schiff GD, Amato MG, Eguale T, Boehne JJ, Wright A, Koppel R, et al. Computerised physician order entry-related medication errors: analysis of reported errors and vulnerability testing of current systems. *BMJ Qual Saf* 2015 Apr 16;24(4):264-271 [FREE Full text] [doi: [10.1136/bmjqs-2014-003555](https://doi.org/10.1136/bmjqs-2014-003555)] [Medline: [25595599](https://pubmed.ncbi.nlm.nih.gov/25595599/)]
55. Staggers N, Jennings BM, Lasome CE. A usability assessment of AHLTA in ambulatory clinics at a military medical center. *Mil Med* 2010 Jul;175(7):518-524. [doi: [10.7205/milmed-d-09-00285](https://doi.org/10.7205/milmed-d-09-00285)] [Medline: [20684457](https://pubmed.ncbi.nlm.nih.gov/20684457/)]
56. Cresswell KM, Mozaffar H, Lee L, Williams R, Sheikh A. Workarounds to hospital electronic prescribing systems: a qualitative study in English hospitals. *BMJ Qual Saf* 2017 Jul;26(7):542-551. [doi: [10.1136/bmjqs-2015-005149](https://doi.org/10.1136/bmjqs-2015-005149)] [Medline: [27129493](https://pubmed.ncbi.nlm.nih.gov/27129493/)]
57. Bhattacharjee S, Murcko AC, Fair MK, Warholak TL. Medication prior authorization from the providers perspective: a prospective observational study. *Res Social Adm Pharm* 2019 Sep;15(9):1138-1144. [doi: [10.1016/j.sapharm.2018.09.019](https://doi.org/10.1016/j.sapharm.2018.09.019)] [Medline: [30279130](https://pubmed.ncbi.nlm.nih.gov/30279130/)]
58. Samaras EA, Real SD, Curtis AM, Meunier TS. Recognizing nurse stakeholder dissonance as a critical determinant of patient safety in new healthcare information technologies. *Work* 2012;41 Suppl 1:1904-1910. [doi: [10.3233/WOR-2012-0406-1904](https://doi.org/10.3233/WOR-2012-0406-1904)] [Medline: [22316994](https://pubmed.ncbi.nlm.nih.gov/22316994/)]
59. Yeung M, Lapinsky S, Granton J, Doran D, Cafazzo J. Examining nursing vital signs documentation workflow: barriers and opportunities in general internal medicine units. *J Clin Nurs* 2012 Apr;21(7-8):975-982. [doi: [10.1111/j.1365-2702.2011.03937.x](https://doi.org/10.1111/j.1365-2702.2011.03937.x)] [Medline: [22243491](https://pubmed.ncbi.nlm.nih.gov/22243491/)]

60. Naughton BD. Medicine authentication technology: a quantitative study of incorrect quarantine, average response times and offline issues in a hospital setting. *BMJ Open* 2019 Feb 19;9(2):e026619 [FREE Full text] [doi: [10.1136/bmjopen-2018-026619](https://doi.org/10.1136/bmjopen-2018-026619)] [Medline: [30782947](https://pubmed.ncbi.nlm.nih.gov/30782947/)]
61. Sockolow P, Bass EJ, Eberle CL, Bowles KH. Homecare nurses' decision-making during admission care planning. *Stud Health Technol Inform* 2016;225:28-32. [Medline: [27332156](https://pubmed.ncbi.nlm.nih.gov/27332156/)]
62. Slight S, Eguale T, Amato M, Seger AC, Whitney DL, Bates DW, et al. The vulnerabilities of computerized physician order entry systems: a qualitative study. *J Am Med Inform Assoc* 2016 Mar;23(2):311-316. [doi: [10.1093/jamia/ocv135](https://doi.org/10.1093/jamia/ocv135)] [Medline: [26568606](https://pubmed.ncbi.nlm.nih.gov/26568606/)]
63. Balka E, Tolar M. Everyday ethical dilemmas arising with electronic record use in primary care. *Stud Health Technol Inform* 2011;169:285-289. [Medline: [21893758](https://pubmed.ncbi.nlm.nih.gov/21893758/)]
64. Vest J, Issel LM, Lee S. Experience of using information systems in public health practice: findings from a qualitative study. *Online J Public Health Inform* 2014 Feb 05;5(3):227 [FREE Full text] [doi: [10.5210/ojphi.v5i3.4847](https://doi.org/10.5210/ojphi.v5i3.4847)] [Medline: [24678380](https://pubmed.ncbi.nlm.nih.gov/24678380/)]
65. Steele Gray C, Barnsley J, Gagnon D, Belzile L, Kenealy T, Shaw J, et al. Using information communication technology in models of integrated community-based primary health care: exploring ICT in the iCOACH study. *Int J Integr Care* 2017 Oct 17;17(5):144. [doi: [10.5334/ijic.3452](https://doi.org/10.5334/ijic.3452)]
66. Østensen E, Bragstad LK, Hardiker NR, Hellesø R. Nurses' information practice in municipal health care-A web-like landscape. *J Clin Nurs* 2019 Jul 23;28(13-14):2706-2716. [doi: [10.1111/jocn.14873](https://doi.org/10.1111/jocn.14873)] [Medline: [30938870](https://pubmed.ncbi.nlm.nih.gov/30938870/)]
67. Remen VM, Grimsmo A. Closing information gaps with shared electronic patient summaries: how much will it matter? *Int J Med Inform* 2011 Nov;80(11):775-781. [doi: [10.1016/j.ijmedinf.2011.08.008](https://doi.org/10.1016/j.ijmedinf.2011.08.008)] [Medline: [21956001](https://pubmed.ncbi.nlm.nih.gov/21956001/)]
68. Eikkey EV, Murphy AR, Reddy MC, Xu H. Designing for privacy management in hospitals: understanding the gap between user activities and IT staff's understandings. *Int J Med Inform* 2015 Dec;84(12):1065-1075. [doi: [10.1016/j.ijmedinf.2015.09.006](https://doi.org/10.1016/j.ijmedinf.2015.09.006)] [Medline: [26467571](https://pubmed.ncbi.nlm.nih.gov/26467571/)]
69. Shemilt K, Morecroft CW, Ford JL, Mackridge AJ, Green C. Inpatient prescribing systems used in NHS Acute Trusts across England: a managerial perspective. *Eur J Hosp Pharm* 2017 Jul 30;24(4):213-217 [FREE Full text] [doi: [10.1136/ejpharm-2016-000905](https://doi.org/10.1136/ejpharm-2016-000905)] [Medline: [31156943](https://pubmed.ncbi.nlm.nih.gov/31156943/)]
70. Mamykina L, Vawdrey DK, Stetson PD, Zheng K, Hripcsak G. Clinical documentation: composition or synthesis? *J Am Med Inform Assoc* 2012 Nov 01;19(6):1025-1031 [FREE Full text] [doi: [10.1136/amiainl-2012-000901](https://doi.org/10.1136/amiainl-2012-000901)] [Medline: [22813762](https://pubmed.ncbi.nlm.nih.gov/22813762/)]
71. Collins SA, Fred M, Wilcox L, Vawdrey DK. Workarounds used by nurses to overcome design constraints of electronic health records. In: Proceedings of the 11th International Congress on Nursing Informatics. 2012 Presented at: International Congress in Nursing Informatics; June 23-27, 2012; Montreal, Canada URL: <http://europepmc.org/abstract/MED/24199061>
72. Stanhope V, Matthews EB. Delivering person-centered care with an electronic health record. *BMC Med Inform Decis Mak* 2019 Aug 22;19(1):168 [FREE Full text] [doi: [10.1186/s12911-019-0897-6](https://doi.org/10.1186/s12911-019-0897-6)] [Medline: [31438960](https://pubmed.ncbi.nlm.nih.gov/31438960/)]
73. Swinglehurst D, Greenhalgh T. Caring for the patient, caring for the record: an ethnographic study of 'back office' work in upholding quality of care in general practice. *BMC Health Serv Res* 2015 Apr 23;15(1):177 [FREE Full text] [doi: [10.1186/s12913-015-0774-7](https://doi.org/10.1186/s12913-015-0774-7)] [Medline: [25907436](https://pubmed.ncbi.nlm.nih.gov/25907436/)]
74. Friedman A, Crosson JC, Howard J, Clark EC, Pellerano M, Karsh B, et al. A typology of electronic health record workarounds in small-to-medium size primary care practices. *J Am Med Inform Assoc* 2014 Feb 01;21(e1):78-83 [FREE Full text] [doi: [10.1136/amiainl-2013-001686](https://doi.org/10.1136/amiainl-2013-001686)] [Medline: [23904322](https://pubmed.ncbi.nlm.nih.gov/23904322/)]
75. Campbell AN, McCarty D, Rieckmann T, McNeely J, Rotrosen J, Wu L, et al. Interpretation and integration of the federal substance use privacy protection rule in integrated health systems: a qualitative analysis. *J Subst Abuse Treat* 2019 Feb;97:41-46 [FREE Full text] [doi: [10.1016/j.jsat.2018.11.005](https://doi.org/10.1016/j.jsat.2018.11.005)] [Medline: [30577898](https://pubmed.ncbi.nlm.nih.gov/30577898/)]
76. Bramble JD, Abbott AA, Fuji KT, Paschal KA, Siracuse MV, Galt K. Patient safety perspectives of providers and nurses: the experience of a rural ambulatory care practice using an EHR with E-prescribing. *J Rural Health* 2013 Mar 12;29(4):383-391. [doi: [10.1111/jrh.12015](https://doi.org/10.1111/jrh.12015)] [Medline: [24088212](https://pubmed.ncbi.nlm.nih.gov/24088212/)]
77. Kandaswamy S, Hettinger A, Hoffman D, Ratwani R, Marquard J. Communication through the electronic health record: frequency and implications of free text orders. *JAMIA Open* 2020 Jul;3(2):154-159 [FREE Full text] [doi: [10.1093/jamiaopen/ooaa020](https://doi.org/10.1093/jamiaopen/ooaa020)] [Medline: [32734153](https://pubmed.ncbi.nlm.nih.gov/32734153/)]
78. Fraczkowski D, Matson J, Lopez KD. Nurse workarounds in the electronic health record: an integrative review. *J Am Med Inform Assoc* 2020 Jul 01;27(7):1149-1165 [FREE Full text] [doi: [10.1093/jamia/ocaa050](https://doi.org/10.1093/jamia/ocaa050)] [Medline: [32651588](https://pubmed.ncbi.nlm.nih.gov/32651588/)]

Abbreviations

EHR: electronic health record

PRISMA: Preferred Reporting Item for Systematic Reviews and Meta-Analyses

SEIPS: Systems Engineering Initiative for Patient Safety

SEWA: Sociotechnical Electronic Health Record Workaround Analysis

Edited by R Kukafka; submitted 22.08.21; peer-reviewed by G Petersson, R Marshall; comments to author 23.10.21; revised version received 13.12.21; accepted 16.12.21; published 15.03.22.

Please cite as:

Blijleven V, Hoxha F, Jaspers M

Workarounds in Electronic Health Record Systems and the Revised Sociotechnical Electronic Health Record Workaround Analysis Framework: Scoping Review

J Med Internet Res 2022;24(3):e33046

URL: <https://www.jmir.org/2022/3/e33046>

doi: [10.2196/33046](https://doi.org/10.2196/33046)

PMID: [35289752](https://pubmed.ncbi.nlm.nih.gov/35289752/)

©Vincent Blijleven, Florian Hoxha, Monique Jaspers. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 15.03.2022. 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.

Review

Effects of a Nurse-Led Telehealth Self-care Promotion Program on the Quality of Life of Community-Dwelling Older Adults: Systematic Review and Meta-analysis

Arkers Kwan Ching Wong¹, PhD; Jonathan Bayuo¹, MPhil; Frances Kam Yuet Wong¹, PhD; Wing Shan Yuen¹, BSc; Athena Yin Lam Lee¹, BSc; Pui King Chang¹, BSc; Jojo Tsz Chui Lai¹, MSc

School of Nursing, The Hong Kong Polytechnic University, Kowloon, Hong Kong

Corresponding Author:

Arkers Kwan Ching Wong, PhD
School of Nursing
The Hong Kong Polytechnic University
Hung Hom Campus
FG 512
Kowloon
Hong Kong
Phone: 852 34003805
Email: arkers.wong@polyu.edu.hk

Abstract

Background: In recent years, telehealth has become a common channel for health care professionals to use to promote health and provide distance care. COVID-19 has further fostered the widespread use of this new technology, which can improve access to care while protecting the community from exposure to infection by direct personal contact, and reduce the time and cost of traveling for both health care users and providers. This is especially true for community-dwelling older adults who have multiple chronic diseases and require frequent hospital visits. Nurses are globally recognized as health care professionals who provide effective community-based care to older adults, facilitating their desire to age in place. However, to date, it is unclear whether the use of telehealth can facilitate their work of promoting self-care to community-dwelling older adults.

Objective: This review aims to summarize findings from randomized controlled trials on the effect of nurse-led telehealth self-care promotion programs compared with the usual on-site or face-to-face services on the quality of life (QoL), self-efficacy, depression, and hospital admissions among community-dwelling older adults.

Methods: A search of 6 major databases was undertaken of relevant studies published from May 2011 to April 2021. Standardized mean differences (SMDs) and their 95% CIs were calculated from postintervention outcomes for continuous data, while the odds ratio was obtained for dichotomous data using the Mantel–Haenszel test.

Results: From 1173 possible publications, 13 trials involving a total of 4097 participants were included in this meta-analysis. Compared with the control groups, the intervention groups of community-dwelling older adults significantly improved in overall QoL (SMD 0.12; 95% CI 0.03 to 0.20; $P=.006$; $I^2=21\%$), self-efficacy (SMD 0.19; 95% CI 0.08 to 0.30; $P<.001$; $I^2=0\%$), and depression level (SMD -0.22 ; 95% CI -0.36 to -0.08 ; $P=.003$; $I^2=89\%$).

Conclusions: This meta-analysis suggests that employing telehealth in nurse-led self-care promotion programs may have a positive impact on older adults, although more studies are needed to strengthen the evidence base, particularly regarding organization and delivery.

Trial Registration: PROSPERO (Prospective International Register of Systematic Reviews) CRD42021257299; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=257299

(*J Med Internet Res* 2022;24(3):e31912) doi:[10.2196/31912](https://doi.org/10.2196/31912)

KEYWORDS

telehealth; meta-analysis; self-care; community-dwelling older adult; nurse

Introduction

Aging populations put tremendous pressure on health and social care systems. Encouraging self-care practices and independent living among older adults has been regarded as one of the best solutions to reduce the demands on costly tertiary and institutional care services [1]. Older adults have the responsibility to make an effort to adopt positive personal health practices according to their own preferences. By adopting such self-care practices, they can maintain their autonomy and independence, and enjoy an improved quality of life (QoL) [2].

Nurses are believed to play the most prominent role in promoting self-care behaviors among older adults [3]. Numerous studies provide evidence of their competence and capability in relation to preventive interventions, including their use of comprehensive and systematic assessments that facilitate early identification of older adults' health complaints [3], their adoption of a holistic caring approach that addresses multiple complaints [4], their capacity to make referrals to other health professionals in a multidisciplinary team if needed [5], and their ability to build a trusting relationship with older adults [6]. However, previous nurse-led self-care promotion interventions relied heavily on a supportive environment that allowed only for face-to-face communication, and so can be difficult to implement in the face of existing barriers in health care institutions, such as time constraints [7], and transportation issues for those with physical or functional limitations [8]. These obstacles can jeopardize the quality of the interventions and the eventual health outcomes and QoL of the older adults in need of care [9]. It is thus better to take those interventions to the community level, including patients' homes, in the hope of obtaining sufficient time, geographical convenience, and greater familiarity and security for the introduction of these preventative measures. Although the new practice may also cost a considerable amount of time and manpower, using telehealth as a solution to delivering care may make possible the realization of this vision of "nurse-led preventive community care for all."

Telehealth refers to the services that bring health care directly to users, generally in their own homes, supported by information and communication technology [10]. It includes but is not limited to social alarms, lifestyle monitoring, remote monitoring of vital signs for diagnosis, and long-distance assessment and education. With the assistance of telecommunication tools such as smartphones, audio or video equipment, or tablets, telehealth changes the geography of health care by introducing person-centered virtual communication contexts, such as videoconferencing, telephone calls, and SMS text messages [11,12]. The benefits of telehealth are evident because from a geographical perspective it enables care to be delivered at a distance and improves access to care under different conditions. For instance, health care providers are able to reach out to older adults who are socially isolated or physically homebound due to diseases, disabilities, or other family roles. It has also helped to minimize the risk of direct transmission of infectious diseases for both health care providers and older adults during the COVID-19 pandemic [13]. Meanwhile, from a psychosocial perspective, it redefines familiar places (eg, the homes of older adults) into spaces of care [11]. Without geographical

restrictions and the associated concerns, both older adults and their health care providers can devote more time and attention on the interventions themselves, resulting in an improvement in the quality of care that is provided. Indeed, these benefits are in accordance with López's [14] view that telehealth is a technological catalyst for the implementation of community-based aging-in-place care systems. It elevates both the access to and quality of nurse-led self-care promotion programs in the community, transforming them into unique and holistic preventative measures that effectively increase the QoL of community-dwelling older adults [15], as well as achieving the goal of relieving the burden on health systems.

Despite the apparent benefits of nurse-led telehealth programs on promoting self-care, reviews are lacking of its impact on the QoL of community-dwelling older adults and on health care systems. Previous reviews have mainly focused on the impact of such programs on caregivers instead of on the older adults themselves [9,16,17]. Some focused on patients with a specific disease or who were in the terminal phase of their life, instead of on a sample representing the general population of community-dwelling older adults [18-20], while others overlooked the leading efforts of nurses in using telecare to promote self-care in the community [21]. Little is therefore known about the effects of nurse-led telecare programs on promoting self-care among community-dwelling older adults.

This study is, to the best of our knowledge, the first systematic review and meta-analysis of randomized controlled trials (RCTs) aimed at summarizing evidence on the effects of nurse-led telehealth self-care programs on community-dwelling older adults compared with the usual on-site or face-to-face care. The particular focus is on the quality of the care that is delivered, as well as on other outcomes including self-efficacy, depression, and hospital admissions. Given the popularity of adopting and sustaining telehealth in promoting self-care during the COVID-19 pandemic and in the near future, the empirical evidence from this study may guide the efforts of policymakers to address challenges in providing services for this large but still overlooked segment of the population.

Methods

Overview

This systematic review and meta-analysis followed the Preferred Reporting Items for Systematic Reviews and Meta-analyses (PRISMA) statement.

Search Strategy

Three investigators (PKC, WSY, and AYLL) independently conducted a literature search using CINAHL, MEDLINE (PubMed), EMBASE (Ovid), PsycINFO (BSCO), Web of Science, the Cochrane Central Register of Controlled Trials, and ClinicalTrials.gov to identify RCTs written in English and published between May 2011 and April 2021. Given the rapidly changing nature of technology and the major changes that have taken place in the field of health care within the past 10 years, the goal was to capture the newest and most relevant evidence related to the use of telehealth as a self-care promotion intervention for community-dwelling older adults. Any

disagreements were resolved by consensus with a fourth author (AKCW).

The following search strategy was used: (telehealth OR telecare OR telemedicine OR gerontechnology OR eHealth OR mHealth OR “mobile health” OR telecommunication OR teleconsultation OR teleconference) AND (self-care OR self-help OR self-management OR “self care” OR “self help” OR “self management”) AND (home OR “home health” OR “home care” OR community) AND (elderly OR aged OR aging OR ageing OR old* OR “older adult*” OR senior OR geriatric OR “older person” OR “elderly person”) AND (nurs* OR nurse-led) AND (random* OR control* OR “usual care”). The online search was supplemented by an extensive hand search of the literature through references identified from retrieved articles. Gray literature such as abstracts, conference proceedings, and editorials was excluded.

Study Selection

The criteria for inclusion in this meta-analysis were: (1) RCT; (2) conducted with adults aged 60 or over and living independently in the community; (3) using telehealth (defined as the use of apps, websites, WhatsApp, SMS text messages, email, social media such as Facebook or Twitter, telephone calls, tablets, software such as Zoom or Microsoft Teams, home remote monitoring devices [reactive or proactive], or any combination of these as a health care delivery channel) as an intervention group component; (4) using a face-to-face or on-site care service as the control group component; (5) intended to empower or promote the self-care of community-dwelling older adults (ie, self-care refers to an activity that individuals undertake on their own behalf to stay fit, maintain good health and functioning, and prevent illness, with or without assistance). Studies were excluded if (1) they focused on cognitively or functionally impaired older adults unable to perform self-care; and (2) they compared 1 or more telecare interventions without a comparison with a control group or with a no intervention control group. As this meta-analysis targeted interventions led by nurses, studies that included an interdisciplinary care team should have had nurses carry out at least 50% of the interventions.

For each article included in the review, data about the participants (country, number of participants, inclusion and exclusion criteria), interventions (components of both intervention and control groups, provider, duration), and outcomes (outcome measures, results) were extracted. These were then compared and analyzed. If the aforementioned data were not available, we contacted the corresponding researcher of the study in question to clarify and request missing information.

Outcomes

The primary outcome of interest was QoL. Secondary outcomes of interest were self-efficacy, depression, and hospital admissions.

Quality Assessment

The potential risk of bias in the included studies was evaluated using Cochrane Collaboration’s tool for assessing the risk of bias according to the Cochrane Handbook for Systematic Reviews of Interventions [22]. This tool was used to assess the quality of the included studies by monitoring 7 domains: random sequence generation, allocation concealment, the blinding of participants and personnel, the blinding of the outcome assessment, incomplete outcome data, selective reporting, and other biases [22]. Three authors (PKC, WSY, and AYLL) independently rated the studies according to the assessment tool. Disagreements were resolved through discussion with a fourth author (AKCW).

Data Synthesis and Statistical Analysis

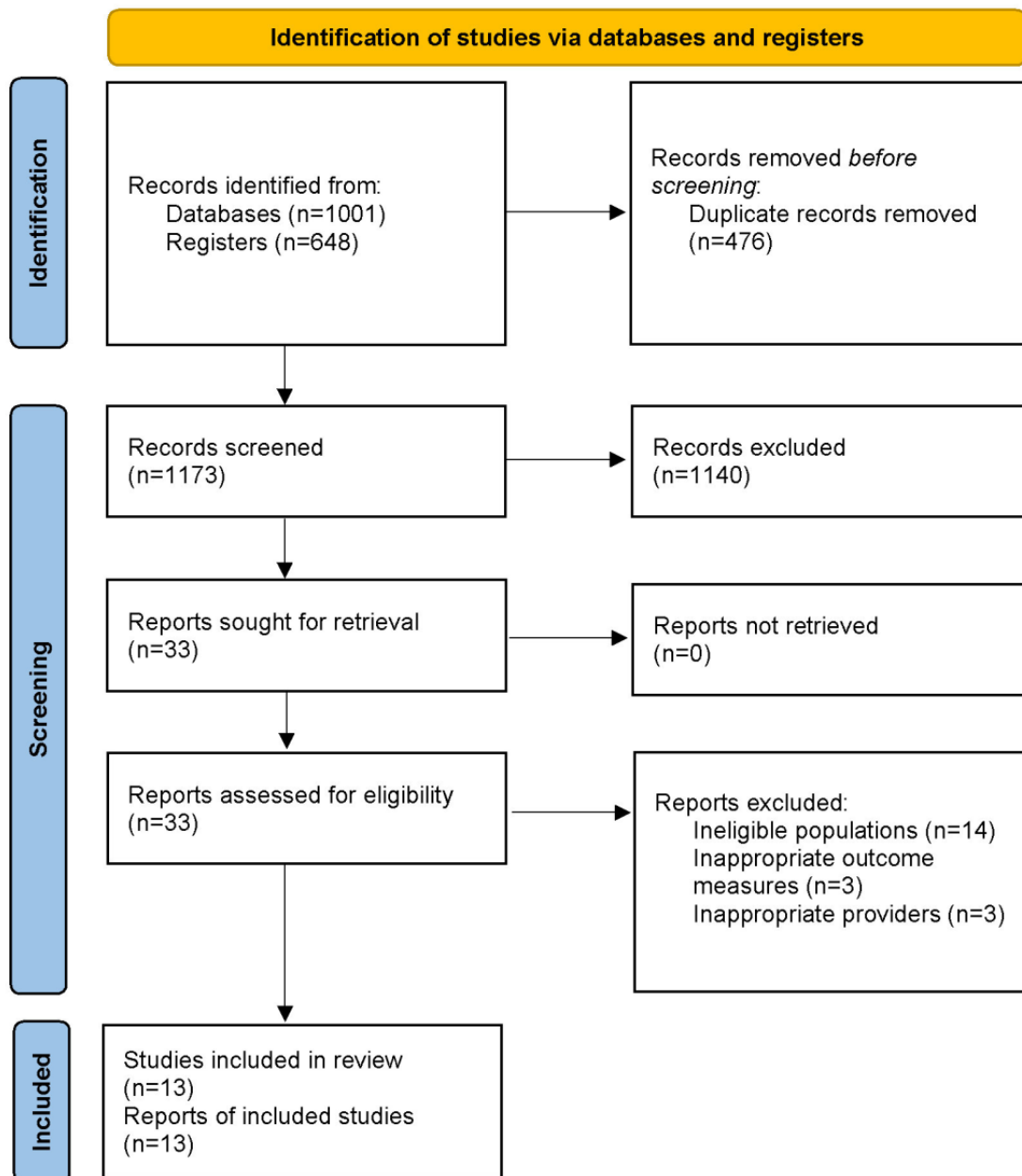
Meta-analyses were conducted using Review Manager (version 5.3). We performed a meta-analysis when a minimum of 2 studies compared the effects of an intervention over the treatment delivered to the control group at the longest follow-up time. Because of the foreseeable complexities and multicomponent nature of nurse-led self-care promotion programs, the research team decided to conduct a random-effects meta-analysis a priori. The accuracy of using this method was tested using a standard χ^2 test and an inconsistency index ($I^2 > 50\%$ or $P < .05$ or both). We planned to run a meta-regression using R (version i386 3.3.2; R Foundation) to explain the between-trial heterogeneity, but because fewer than 10 trials were included, such an approach was not possible [22]. The standardized mean differences (SMDs) and their 95% CIs were calculated from the postintervention outcomes for continuous data, while the odds ratio (ORs) was obtained for dichotomous data by using the Mantel–Haenszel test. The SMD effect sizes were considered small, moderate, and large when the value was < 0.4 , $0.4-0.7$, and > 0.7 , respectively [22]. Pooled ORs (95% CI) were calculated and a 2-sided P -value < 0.1 was adopted to indicate statistical significance [22]. Where a sensitivity analysis was required, the analysis was repeated but with the exclusion of studies with a low study quality/high risk of bias, or lacking a thorough explanation of the timeframe of the reported outcome, the study design, or participant characteristics. Publication bias was checked using a visual inspection of funnel plots [22] and calculated using the Egger bias test [23].

Results

Search Outcomes

We identified 1173 publications in our literature search after the removal of duplicates. Of these, 1140 publications were excluded based on an evaluation of the title and the brief abstract. The remaining 33 publications were assessed for eligibility, and 13 were included in our meta-analysis [24-36]. The most common reason for excluding a study was that the population studied was ineligible ($n=14$; Figure 1). A consensus between 2 independent reviewers was reached in 94% (31/33) of the publications.

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



Quality Assessment and Publication Bias

Overall, the quality of the included RCTs was high, except in the aspects of the blinding of participants and personnel, and allocation concealment (Figure 2). Two studies were deemed to be of poor methodological quality [27,31], 2 of fair quality

[26,32], and the remainder of high quality [24,25,28-30,33-36]. However, 5 studies [26,29-32] were deemed to be at an unclear risk of additional biases, through possible failures in randomization, no mention of baseline differences, and concerns over the power of the study. A summary of the risks of bias of included studies is shown in Table 1.

Figure 2. Risk of bias table.

Study	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Chau et al [24]	?	?	-	-	+	+	-
Chow and Wong [25]	+	+	+	+	+	+	+
De San Miguel et al [26]	+	+	?	?	+	+	?
Finkelstein et al [27]	?	?	?	-	-	+	-
Finlayson et al [28]	+	+	+	+	+	+	+
Gellis et al [29]	+	?	+	+	+	-	?
Jolly et al [30]	+	+	-	+	+	+	?
Kazawa et al [31]	-	?	?	?	+	+	?
Kleinpell et al [32]	+	?	+	+	+	+	?
Oksman et al [33]	+	+	+	+	+	+	+
Pecina et al [34]	?	?	?	+	+	+	?
Takahashi et al [35]	+	+	+	+	+	+	+
Wong et al [36]	+	+	+	+	+	+	+

Table 1. Risk of bias in the included studies.

Study	Random sequence generation	Allocation concealment	Blinding of participants and personnel	Blinding of outcome assessment	Incomplete outcome data	Selective reporting	Other biases
Chau et al [24]	Unclear	Unclear	High	High	Low	Low	High
Chow and Wong [25]	Low	Low	Low	Low	Low	Low	Low
De San Miguel et al [26]	Low	Low	Unclear	Unclear	Low	Low	Unclear
Finkelstein et al [27]	Unclear	Unclear	Unclear	High	High	Low	High
Finlayson et al [28]	Low	Low	Low	Low	Low	Low	Low
Gellis et al [29]	Low	Unclear	Low	Low	Low	High	Unclear
Jolly et al [30]	Low	Low	High	Low	Low	Low	Unclear
Kazawa et al [31]	High	Unclear	Unclear	Unclear	Low	Low	Unclear
Kleinpell et al [32]	Low	Unclear	Low	Low	Low	Low	Unclear
Oksman et al [33]	Low	Low	Low	Low	Low	Low	Low
Pecina et al [34]	Unclear	Unclear	High	Unclear	Low	Low	High
Takahashi et al [35]	Low	Low	Low	Low	Low	Low	Low
Wong et al [36]	Low	Low	Low	Low	Low	Low	Low

Characteristics of the Studies and Participants

Among the 13 publications, 4097 older adults were included in the meta-analysis, with 2096 older adults in intervention groups and 2001 older adults serving as controls [24-36]. The mean age of the entire sample was 73.2 (SD 4.5) years and females made up 68% (2669/3925) of the samples. The telecommunication tools that were adopted in these studies included telephones [25,28,30,33,36], home telemonitoring

devices [24,26,27,29,32,34,35], and videoconferencing software or apps [27,31,35]. A total of 3 studies had nurse case managers providing telehealth services to the participants [25,29,36], another 3 studies had advanced practice nurses delivering the intervention [28,32,34], while the remainder involved registered nurses or community nurses [24,26,27,30,31,33,35]. The duration of the interventions varied from 4 weeks to 48 weeks, with a median of 24 weeks. The characteristics of the included studies are summarized in Table 2.

Table 2. Characteristics of the included studies.

Study	Country	Number of participants	Inclusion/exclusion criteria	Intervention components	Control group	Providers	Duration	Outcome measures	Results
Chau et al [24]	Hong Kong	N=40 (I ^a : 22, C ^b : 18)	<ul style="list-style-type: none"> Inclusion: aged 60 and older, with moderate or severe COPD^c according to the classification of the Global Initiative of Obstructive Lung Disease, admitted to hospital at least once for exacerbation during the previous year. Exclusion: unable to communicate, had impaired cognitive function, illiterate, had hearing problems, or unable to operate the telecare device. 	Home visits with education on self-care and symptom management techniques; a device kit (a specially designed mobile phone, a respiratory rate sensor, and a pulse oximeter), which is used for participants' self-monitoring of oxygen saturation, pulse rate, and respiration rate	Only home visits with education on self-care and symptom management techniques	Community nurse	8 weeks	<ul style="list-style-type: none"> Hospital admissions 	OR ^d 2.33 (95% CI 0.51 to 10.78)
Chow and Wong [25]	Hong Kong	N=281 (I: 96, C: 185)	<ul style="list-style-type: none"> Inclusion: aged 65 and older; admitted with a medical diagnosis related to chronic respiratory, cardiac, type 2 diabetes mellitus, or renal diseases; able to speak Cantonese and to communicate; resident in the hospital service area; and able to be contacted by telephone after discharge. Exclusion: identified as having cognitive problems, Mini-Mental State Examination score of <20; discharged to institutional care; followed by a designated disease management program after discharge; unable to communicate; and terminally ill. 	Telephone calls, comprehensive assessment based on the OMAHA system, analysis of self-care barriers, development of mutual self-care goals, evaluation of interventions	Home visits, social calls	Nurse case managers, senior year nursing students	4 weeks	<ul style="list-style-type: none"> Physical component of QoL^e Mental component of QoL Self-efficacy 	<ul style="list-style-type: none"> SMD^f 0.23 (95% CI -0.01 to 0.48) SMD 0.03 (95% CI -0.22 to 0.56) SMD 0.17 (95% CI -0.07 to 0.42)
De San Miguel et al [26]	Australia	N=71 (I: 36, C: 35)	<ul style="list-style-type: none"> Inclusion: Silver Chain clients with a diagnosis of COPD, receiving domiciliary oxygen, able to speak English, living in the metropolitan area. Exclusion: diagnosed with dementia, receiving palliative care, did not have a telephone landline, unable to use the telehealth equipment because of cognitive or physical impairment. 	Telehealth equipment (Health-HUB), daily measurements, recording and monitoring of vital signs, assessment of general state of health, home visits, educational book about COPD, telehealth instruction manual, telephone calls, provision of support/advice/recommendations	Home visits, education book about COPD	Telehealth nurse	24 weeks	<ul style="list-style-type: none"> Hospital admissions 	OR 0.28 (95% CI 0.10 to 0.76)

Study	Country	Number of participants	Inclusion/exclusion criteria	Intervention components	Control group	Providers	Duration	Outcome measures	Results
Finkelstein et al [27]	United States	N=84 (I: 40, C: 44)	<ul style="list-style-type: none"> Inclusion: aged 60 and older, managing 1 or more chronic diseases, not receiving Medicare home health benefits but had functional limitations, able to manipulate a computer keyboard or a mouse, and had a broadband connection available in their area. Exclusion: not mentioned. 	Home telehealth program using the VALUE workstation, videoconferences, electronic messages, ordering of health-related and home care services, access to health-related information, general access to the internet, physiological monitoring devices	Usual care	Telehealth nurse	8.5 weeks	<ul style="list-style-type: none"> Hospital admissions 	OR 0.41 (95% CI 0.15 to 1.14)
Finlayson et al [28]	Australia	N=222 (I: 111, C: 111)	<ul style="list-style-type: none"> Inclusion: aged 65 and older; admitted with a medical condition; had at least one risk factor for readmission (aged 75 or older, admitted to a hospital more than once in the previous 6 months, multiple comorbidities, living alone, poor social support, poor self-rating of health, functional impairment, or a history of depression). Exclusion: requires home oxygen, dependent on a wheelchair or unable to walk independently for 3 m, lives in a nursing home, presence of a cognitive deficit or progressive neurological disease. 	Tailored exercise program, in-home visits, telephone follow-ups, reinforcement and further explanation of the exercise program, advice and support to the caregiver	Usual care, exercise program without regular telephone follow-ups	Advanced practiced nurse, exercise physiologist	24 weeks	<ul style="list-style-type: none"> Hospital admissions 	OR 0.40 (95% CI 0.17 to 0.92)
Gellis et al [29]	United States	N=94 (I: 48, C: 46)		The Honeywell "HomMed" Health Monitoring System for daily monitoring of weight, non-invasive blood pressure, pulse, oxygen saturation, and temperature; further evaluation of abnormal readings by telehealth nurse, education and counseling on disease, self-care activities, and symptom management strategies	Usual care, education	Home-care telehealth nurse manager, registered homecare nurses	12 weeks	<ul style="list-style-type: none"> Mental component of QoL Depression 	<ul style="list-style-type: none"> SMD 0.45 (95% CI 0.04 to 0.86)^g SMD -1.10 (95% CI -1.53 to -0.66)^g

Study	Country	Number of participants	Inclusion/exclusion criteria	Intervention components	Control group	Providers	Duration	Outcome measures	Results
			<ul style="list-style-type: none"> Inclusion: aged 65 and older, diagnosed with heart failure or COPD, experienced frequent health care encounters (ie, hospitalized twice in the last 6 months or seen at least twice in the emergency room in the past 2 months), required 3 or more home visits per week, consented to participate in the program with random assignment, expressed a willingness to learn how to use the telehealth monitoring system. Exclusion: unable to learn to use the HomMED telehealth device due to physical disability, cognitively impaired based on a medical chart diagnosis and had no caregiver, exhibited behavioral/problems (eg, aggression, agitation, delirium, paranoia) that interfered with learning how to use the HomMED telehealth device and communicating with the telehealth nurse. 						
Jolly et al [30]	UK	N=516 (I: 239, C: 277)	<ul style="list-style-type: none"> Inclusion: has respiratory symptoms consistent with COPD, reported mild dyspnea at the baseline assessment, had a forced expiratory volume in 1 second/forced vital capacity score of <0.7 after postbronchodilator spirometry (consistent with current UK guidelines) at the baseline assessment. Exclusion: considered by doctors to be inappropriate for inclusion (eg, for having a terminal disease or a severe psychiatric disorder) 	Telephone health coaching with supporting written documents, a pedometer, and a self-monitoring diary	Usual care with a standard information leaflet about the self-management of COPD	Nurse	24 weeks	<ul style="list-style-type: none"> QoL Self-efficacy Depression 	<ul style="list-style-type: none"> SMD 0.18 (95% CI 0.00 to 0.36)^g SMD 0.23 (95% CI 0.06 to 0.41)^g SMD -0.15 (95% CI -0.33 to 0.03)
Kazawa et al [31]	Japan	N=32 (I: 17, C: 15)					24 weeks		

Study	Country	Number of participants	Inclusion/exclusion criteria	Intervention components	Control group	Providers	Duration	Outcome measures	Results
			<ul style="list-style-type: none"> Inclusion: had a proteinuria level of $\geq 2+$ or a proteinuria level of 1+ and a hemoglobin A1c level of $\geq 7.0\%$ (or a fasting blood sugar level of ≥ 130 mg/dL) at a health check conducted in 2013, and diagnosed with type 2 diabetes mellitus. Exclusion: has type 1 diabetes mellitus or gestational diabetes, had initiated dialysis, scheduled for renal transplantation in the near future, undergoing treatment for cancer, has a terminal illness, has cognitive impairment, or has a mental disorder. 	Distance interviews via a tablet with a featured app (delivered to the participants by postal mail), a guidebook, a self-monitoring notebook, and foot care monofilament	Direct face-to-face interviews and intermittent telephone calls	Nurse trained in disease management		<ul style="list-style-type: none"> Systolic blood pressure (mmHg) Diastolic blood pressure (mmHg) BMI QoL Self-efficacy 	<ul style="list-style-type: none"> SMD 6.50 (95% CI -1.44 to 14.44) SMD 0.80 (95% CI -4.02 to 5.62) SMD 3.50 (95% CI 0.55 to 6.45)^g SMD 0.68 (95% CI -0.15 to 1.51) SMD -0.42 (95% CI -1.23 to 0.39)
Kleinpell et al [32]	United States	N=206 (I: 134, C: 72)	<ul style="list-style-type: none"> Inclusion: aged ≥ 65 at high risk for postoperative complications; documented history of congestive heart failure; New York Heart Association functional classification of III or IV; ejection fraction of $\leq 40\%$; a history of atrial fibrillation; postdischarge complications of myocardial infarction, arrhythmias requiring treatment, reoperation, cardiac arrest, wound dehiscence, a positive wound culture; ICU^h stay of >2 days, mechanical ventilation for >2 days; or failure to meet clinical pathway discharge goals by postoperative day 5. Exclusion: not mentioned. 	Home telemonitoring twice daily of vital signs including heart rate, blood pressure, and pulse oximetry, and daily monitoring of weight, focused reinforcement of the discharge plan	No intervention	Advanced practice nurse	4 weeks	<ul style="list-style-type: none"> QoL Hospital admissions 	<ul style="list-style-type: none"> SMD -0.13 (95% CI -0.14 to 0.16) OR 0.70 (95% CI 0.32 to 1.54)
Oksman et al [33]	Finland	N=1570 (I: 970, C: 470)			Routine social and health care	Certified nurses and public health nurses	48 weeks	<ul style="list-style-type: none"> QoL 	<ul style="list-style-type: none"> SMD 0.12 (95% CI 0.01 to 0.23)^g

Study	Country	Number of participants	Inclusion/exclusion criteria	Intervention components	Control group	Providers	Duration	Outcome measures	Results
			<ul style="list-style-type: none"> Inclusion: has a glycosylated hemoglobin (hemoglobin A1c) level of >7, or a total cholesterol level of >4.5, or a low-density lipoprotein level of >2.3 for the previous 6 months, identified by a research nurse as being eligible for coaching. Exclusion: classified as ineligible by primary care physician, unable to co-operate or participate in health coaching, major elective surgery planned within 6 months, history of major surgery within the past 2 years, life expectancy <1 year, pregnancy. 	Individual health coaching by telephone, in addition to routine social and health care, including 8 key recommendations developed by Pfizer Health Solutions: (1) know how and when to call for help, (2) learn about the condition and set goals, (3) take medicines correctly, (4) get recommended tests and services, (5) act to keep the condition well, (6) make lifestyle changes and reduce risk, (7) build on strengths and overcome obstacles, and (8) follow-up with specialists and appointments					
Pecina et al [34]	United States	N=166 (I: 77, C: 89)	<ul style="list-style-type: none"> Inclusion: aged 60 years and older with an ERAⁱ score in the highest decile. The ERA score is a composite score of previous hospitalizations, age, race, and presence of chronic disease. A high ERA score indicates an increased risk of hospitalization and emergency department visits. Exclusion: unable or unwilling to use the monitoring equipment, or if there was a concern about undiagnosed dementia after a mental status test. 	Telemonitoring of biometric data using an Intel Health Guide device, questionnaires on symptoms, video-conference visits	Usual care	Geriatric nurse practitioner	48 weeks	<ul style="list-style-type: none"> QoL Physical component of QoL Mental component of QoL Depression 	<ul style="list-style-type: none"> SMD 0.11 (95% CI -0.20 to 0.41) SMD -0.35 (95% CI -0.65 to -0.04) SMD -0.02 (95% CI -0.33 to 0.28) SMD 0.00 (95% CI -0.31 to 0.31)
Takahashi et al [35]	United States	N=205 (I: 102, C: 103)		Telemonitoring device (Intel Health Guide; Intel-GE) with real-time videoconferencing capability and peripheral measures (scales, blood pressure cuff, glucometer, pulse oximeter, and peak flow data)	Usual care	Registered nurse	48 weeks	<ul style="list-style-type: none"> Hospital admissions 	<ul style="list-style-type: none"> OR 0.93 (95% CI 0.06 to 14.94)

Study	Country	Number of participants	Inclusion/exclusion criteria	Intervention components	Control group	Providers	Duration	Outcome measures	Results
			<ul style="list-style-type: none"> Inclusion: aged ≥ 60; enrolled in the Employee and Community Health program primary care panel and whose ERA score exceeded 15. The ERA is an electronic database used to assess patient risk for hospitalizations or emergency department visits based on administrative data on age, sex, previous hospitalizations, and comorbid conditions (stroke, dementia, heart disease, diabetes mellitus, and chronic obstructive pulmonary disease). Exclusion: lives in a nursing home, has a clinical diagnosis of dementia or scored 29 or less in the short test of mental status, unable to use the telemonitoring system (ie, because of visual impairment or an inability to use the device). 						
Wong et al [36]	Hong Kong	N=610 (I: 204, C: 406)	<ul style="list-style-type: none"> Inclusion: admitted with a primary diagnosis related to a respiratory, diabetic, cardiac, or renal condition; Mini-Mental State Examination score of >20; able to speak Cantonese; lives within the service area; can be contacted by phone. Exclusion: discharged to an assisted care facility, being followed up by an immediate designated disease management program after discharge, unable to communicate, discharged for end-of-life care. 	Telephone calls, comprehensive assessment based on the OMAHA system, develop mutual self-care goals, evaluate interventions	Home visits, placebo calls (ie, social calls)	Nurse case managers, trained nursing students	4 weeks	<ul style="list-style-type: none"> Self-efficacy Hospital admissions 	<ul style="list-style-type: none"> SMD 0.19 (95% CI 0.02 to 0.36)^g OR 0.84 (95% CI 0.56 to 1.26)

^aI: intervention group.

^bC: control group.

^cCOPD: chronic obstructive pulmonary disease.

^dOR: odds ratio.

^eQoL: quality of life.

^fSMD: standardized mean difference.

^gStatistically significant.

^hICU: intensive care unit.

ⁱERA: Elderly Risk Assessment.

Quantitative Synthesis

Quality of Life

Overview

A total of 5 of the 13 (38%) studies were RCTs that compared the effects of a nurse-led telehealth self-care promotion program with the usual care on the QoL of community-dwelling older adults [30-34]. The pooled SMD in the overall score for QoL was significantly different (SMD 0.12; 95% CI 0.03 to 0.20; $P=.006$; $I^2=21\%$), with the participants in the intervention group having a better QoL than those in the control group.

Physical Component of Quality of Life

Two studies assessed the physical component of QoL by using the Medical Outcomes Study Short Form Survey [25,34]. Pooled

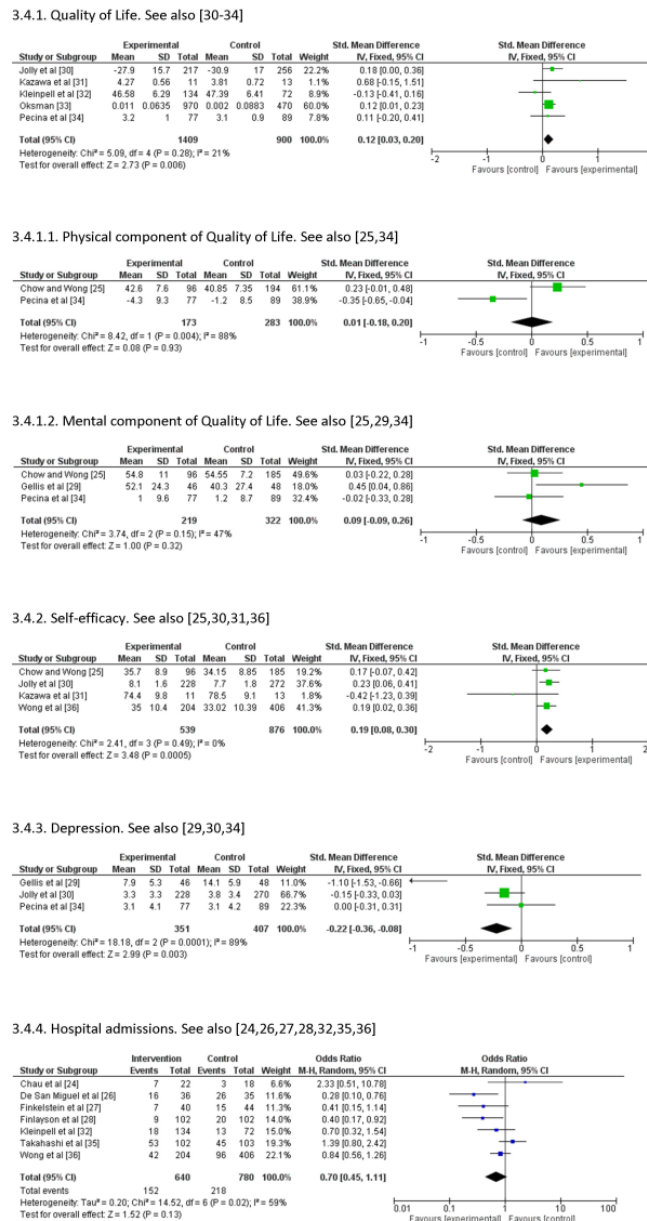
analyses showed that a telehealth self-care promotion program did not lead to an improvement in physical component of QoL over the usual care (SMD 0.01; 95% CI -0.18 to 0.20; $P=.93$), with high heterogeneity ($\chi_1^2=8.42$; $I^2=88\%$; $P=.004$).

Mental Component of Quality of Life

As shown in Figure 3, the telehealth self-care promotion program did not significantly improve the mental component of QoL when compared with the usual care in the 3 studies (SMD 0.09; 95% CI -0.09 to 0.26; $P=.32$) [25,29,34]. The I^2 statistics reflected moderate heterogeneity among the studies ($\chi_2^2=3.74$; $I^2=47\%$; $P=.15$).

None of these outcomes showed evidence of publication bias as revealed by a visual inspection of funnel plots or the P -values of the Egger test ($P>.05$).

Figure 3. Forest plots showing the effects of nurse-led telehealth self-care promotion programmes on different outcomes.



Self-efficacy

Four studies assessed self-efficacy [25,30,31,36], of which 2 found that the telehealth self-care promotion program had a significantly beneficial effect over the usual face-to-face care [30,36]. The pooled SMD in the overall score for self-efficacy was significantly different (SMD 0.19; 95% CI 0.08 to 0.30; $P < .001$). No evidence of heterogeneity ($\chi^2 = 2.41$; $I^2 = 0\%$; $P = .49$) was found and there was no sign of publication bias ($P = .71$).

Depression

The pooled SMD in the overall score for depression was significantly different (SMD -0.22; 95% CI -0.36 to -0.08; $P = .003$) in a meta-analysis of 3/13 studies (23%) [29,30,34]. High heterogeneity ($\chi^2 = 18.2$; $I^2 = 89\%$; $P = .009$) was indicated, but no sign of publication bias ($P = .50$) was found.

Hospital Admissions

Hospital admissions were reported as the outcome in 7/13 studies (54%), with 1420 participants [24,26-28,32,35,36]. Moderate heterogeneity was found among these studies ($\chi^2 = 14.5$; $I^2 = 59\%$; $P = .02$). The number of hospital admissions in the telehealth group was 152 out of 640 (23.8%) and in the usual face-to-face group of participants was 218 out of 780 (27.9%). No significant difference was found between the groups in the number of hospital admissions (OR 0.70, 95% CI 0.45-1.11; $P = .13$).

The forest plots of all outcomes are presented in Figure 3.

Discussion

Principal Findings

In this review an attempt is made to summarize the evidence to ascertain the effects of nurse-led telehealth self-care programs

for community-dwelling older adults in terms of QoL, self-efficacy, levels of depression, and hospital admissions. Overall, the findings of this review suggest that nurse-led telehealth programs may improve the QoL, self-efficacy, and depression levels of community-dwelling older adults when compared with the usual face-to-face care. However, no significant differences across groups were noted in hospital admissions. Although the studies seem limited in some respects, the findings of this review offer insights into the potential effectiveness of employing assistive technologies in community-based health and social care programs and on how these technologies affect the daily life of older adults, although more studies are needed to strengthen the evidence base, particularly in the aspects of organization and delivery.

Undoubtedly, the emergence of COVID-19 has led to a great global need to restructure health and social care services across patient groups, particularly regarding innovative strategies that actively support clients and their family caregivers even at a distance [37]. For community-dwelling older adults who require continuous monitoring, professional support at a distance may be an invaluable add-on to promote self-care practices. As highlighted in this review, nurse-led programs of care may lead to improvements in QoL, self-efficacy, and depression, making it a form of professional support that merits consideration. Even in studies where statistically significant findings were not observed, improvements in health outcomes such as QoL and self-efficacy were noted [31], as well as improved self-management practices [30]. A similar pattern of results was reported among persons living with cancer [38] and type 2 diabetes mellitus [39] who received nurse-led services. Taken together, the findings seem to suggest that well-designed nurse-led services delivered by trained staff may be a promising program of care that can complement and extend existing services from the health care facility to the home/community. There is, however, a need to standardize the contents and dosages of nurse-led services tailored to varied patient groups and to test these using large-scale, well-designed RCTs to strengthen the evidence base regarding their effectiveness in improving other health outcomes. In addition, a process evaluation following implementation may clarify contextual factors that can hinder or facilitate the delivery of the nurse-led programs of care and offer greater explanatory power regarding the impact of a program.

The telehealth component of the nurse-led programs of care mainly comprised structured telephone follow-ups that played an essential role in delivering education, advocacy, and coaching/behavioral change strategies. In addition, the use of customized telehealth monitoring systems installed in the homes of participants or utilized as wearable tracking devices was noted in 7 studies [24,26,27,29,32,34,35]. Evidently, as the demand for access to health care grows along with the aging population, the real-time monitoring of various physiological parameters will become a significant component of health care. Telehealth, which represents the intersection of health and technology, offers unique opportunities to deliver personalized care. The findings of this study should enable researchers and policymakers to better understand the various technologies and their effectiveness. With this understanding, they can better

advise older adults on how to improve their QoL and self-efficacy and reduce their depression using appropriate assistive technologies. Besides, governments should recognize and promote the use of new technologies and the positive impact of these technologies on society, health care, and the QoL of older adults. This is because the use of these technologies not only improves the QoL of older adults but also has a positive impact on the health care system by potentially reducing health care service utilization.

Another key finding in this review is the effect of the nurse-led telehealth services on hospital admissions, which was noted to be statistically insignificant across groups. In previous studies evaluating the effects of nurse-led programs of care, the findings regarding hospital admissions were mixed. A recent integrative review that included 9 studies concluded that there is no clear evidence that community nurse-led services for older persons reduced hospital readmissions [40]. A similar finding was reported by studies involving other patient groups such as children discharged from hospital [41] and persons with heart failure [42]. By contrast, in a nurse-led program of care that focused on delivering a 4-week self-help and empowerment program for older adults living with chronic diseases, a significantly lower admission rate was observed for the intervention group compared with the control group within 84 days of an index admission [25]. Similar findings on nurse-led interventions leading to lower readmission rates have also been reported among persons with heart failure [43,44]. Although the mixed findings may be related to the nature of the interventions, the context of their delivery, or the timeline for the endpoint outcome assessment, it is likely that the intensity of the needs of the individual patients contributed to the hospitalization rates that were observed. In addition, the limitation regarding sample size across studies might make it difficult to draw conclusions. Thus, future studies are needed to address this concern/limitation to enable stronger conclusions to be drawn.

Limitations

This meta-analysis has a few limitations. First, moderate to high heterogeneity was identified among studies that measured depression, the physical and mental components of QoL, and hospital admissions, because only 2 or 3 studies were available on these outcomes. While these studies also varied in terms of duration, content, length of follow-up, and telecommunication tool used in the programs, it was difficult to control for these differences by conducting a sensitivity analysis or a meta-regression (because there were fewer than 10 studies). Second, this study did not exclude disease-specific or transitional self-management programs that were provided by hospital-based health care professionals. Although these programs were also intended to promote self-care and health among older adults, they emphasized disease-specific skill-based training that may have been different from that in the other included studies. Participants might also have been more aware of their health after hospitalization and more willing to adhere to the recommendations of health care professionals, which led to the deviations in the results of the meta-analysis. A subgroup analysis, however, did not reveal differences between studies that focused on older adults with a specific disease and a general

older population. Third, the outcome measures chosen in this study relied on subjective reports from the participants. Future RCTs may benefit from incorporating objective measurements of self-care behavior such as frequency of exercise, BMI, and the pursuit of a healthy diet.

Conclusions

This meta-analysis of 13 RCTs revealed that nurse-led telehealth self-care promotion programs may effectively improve quality of care and self-efficacy, and alleviate depression among

community-dwelling older adults. Despite the methodological limitations of the studies, including variations in the included samples, the intervention content, and the duration across studies, these results may be crucial for policymakers and health care providers to refer to when planning and designing an effective self-care health promotion program to empower older adults to take an active role in taking care of their health, be responsive to their care needs, and eventually to stay in the community with optimal well-being through the use of telehealth.

Acknowledgments

This work was supported by a grant from the Nethersole Institute of Continuing Holistic Health Education (NICHE) (reference number P0034141). The funders had no role in the design of the study; in the collection, analysis, and interpretation of the data; in the writing of the report; or in the decision to submit the article for publication.

Authors' Contributions

AKCW was responsible for conceptualization, methodology, formal analysis, writing—original draft, supervision, funding acquisition. JB took care of methodology, writing—original draft, writing—review and editing. FKYW was responsible for writing—review and editing, funding acquisition. WSY played an active role in validation and investigation. AYLL and PKC were involved in investigation. JTCL was responsible for investigation, writing—original draft.

Conflicts of Interest

None declared.

References

1. Mountain G, Windle G, Hind D, Walters S, Keertharuth A, Chatters R, et al. A preventative lifestyle intervention for older adults (lifestyle matters): a randomised controlled trial. *Age Ageing* 2017 Jul 01;46(4):627-634 [FREE Full text] [doi: [10.1093/ageing/afx021](https://doi.org/10.1093/ageing/afx021)] [Medline: [28338849](https://pubmed.ncbi.nlm.nih.gov/28338849/)]
2. Active ageing: A policy framework. World Health Organization. Geneva, Switzerland: World Health Organization; 2002. URL: <https://extranet.who.int/agefriendlyworld/wp-content/uploads/2014/06/WHO-Active-Ageing-Framework.pdf> [accessed 2021-06-23]
3. Kempainen V, Tossavainen K, Turunen H. Nurses' roles in health promotion practice: an integrative review. *Health Promot Int* 2013 Dec;28(4):490-501. [doi: [10.1093/heapro/das034](https://doi.org/10.1093/heapro/das034)] [Medline: [22888155](https://pubmed.ncbi.nlm.nih.gov/22888155/)]
4. Borglin G, Jakobsson U, Edberg A, Hallberg IR. Self-reported health complaints and their prediction of overall and health-related quality of life among elderly people. *Int J Nurs Stud* 2005 Feb;42(2):147-158. [doi: [10.1016/j.ijnurstu.2004.06.003](https://doi.org/10.1016/j.ijnurstu.2004.06.003)] [Medline: [15680613](https://pubmed.ncbi.nlm.nih.gov/15680613/)]
5. Iliffe S. Community-based interventions for older people with complex needs: time to think again? *Age Ageing* 2016 Jan;45(1):2-3. [doi: [10.1093/ageing/afv185](https://doi.org/10.1093/ageing/afv185)] [Medline: [26764386](https://pubmed.ncbi.nlm.nih.gov/26764386/)]
6. Hupcey JE, Miller J. Community dwelling adults' perception of interpersonal trust vs. trust in health care providers. *J Clin Nurs* 2006 Sep;15(9):1132-1139. [doi: [10.1111/j.1365-2702.2006.01386.x](https://doi.org/10.1111/j.1365-2702.2006.01386.x)] [Medline: [16911054](https://pubmed.ncbi.nlm.nih.gov/16911054/)]
7. Johansson H, Weinehall L, Emmelin M. "If we only got a chance." Barriers to and possibilities for a more health-promoting health service. *J Multidiscip Healthc* 2009 Dec 22;3:1-9 [FREE Full text] [doi: [10.2147/jmdh.s8104](https://doi.org/10.2147/jmdh.s8104)] [Medline: [21197350](https://pubmed.ncbi.nlm.nih.gov/21197350/)]
8. Shi L, Starfield B, Politzer R, Regan J. Primary care, self-rated health, and reductions in social disparities in health. *Health Serv Res* 2002 Jun;37(3):529-550 [FREE Full text] [doi: [10.1111/1475-6773.t01-1-00036](https://doi.org/10.1111/1475-6773.t01-1-00036)] [Medline: [12132594](https://pubmed.ncbi.nlm.nih.gov/12132594/)]
9. Williams K, Blyler D, Vidoni ED, Shaw C, Wurth J, Seabold D, et al. A randomized trial using telehealth technology to link caregivers with dementia care experts for in-home caregiving support: FamTechCare protocol. *Res Nurs Health* 2018 Jun;41(3):219-227 [FREE Full text] [doi: [10.1002/nur.21869](https://doi.org/10.1002/nur.21869)] [Medline: [29504666](https://pubmed.ncbi.nlm.nih.gov/29504666/)]
10. Audit Commission. Implementing Telecare - The Learning Exchange. London, UK: Audit Commission; 2004. URL: <https://tinyurl.com/5n7amyde> [accessed 2022-02-26]
11. Oudshoorn N. Telecare Technologies and the Transformation of Healthcare. London, UK: Palgrave Macmillan; 2011.
12. Vega D. Automated interoperability testing of healthcare information systems. In: *Advances in Computers*. Amsterdam, The Netherlands: Elsevier; 2012:213-276.
13. 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](https://doi.org/10.1186/s12889-020-09301-4)] [Medline: [32738884](https://pubmed.ncbi.nlm.nih.gov/32738884/)]

14. López D. *New Technologies and Emerging Spaces of Care*. London, UK: Routledge; 2016:55-72.
15. Chou C, Chang C, Lee T, Chou H, Mills ME. Technology acceptance and quality of life of the elderly in a telecare program. *Comput Inform Nurs* 2013 Jul;31(7):335-342. [doi: [10.1097/NXN.0b013e318295e5ce](https://doi.org/10.1097/NXN.0b013e318295e5ce)] [Medline: [23728446](https://pubmed.ncbi.nlm.nih.gov/23728446/)]
16. Chi N, Demiris G. A systematic review of telehealth tools and interventions to support family caregivers. *J Telemed Telecare* 2015 Jan;21(1):37-44 [FREE Full text] [doi: [10.1177/1357633X14562734](https://doi.org/10.1177/1357633X14562734)] [Medline: [25475220](https://pubmed.ncbi.nlm.nih.gov/25475220/)]
17. Marzorati C, Renzi C, Russell-Edu SW, Pravettoni G. Telemedicine use among caregivers of cancer patients: systematic review. *J Med Internet Res* 2018 Jun 18;20(6):e223 [FREE Full text] [doi: [10.2196/jmir.9812](https://doi.org/10.2196/jmir.9812)] [Medline: [29914858](https://pubmed.ncbi.nlm.nih.gov/29914858/)]
18. So CF, Chung JW. Telehealth for diabetes self-management in primary healthcare: A systematic review and meta-analysis. *J Telemed Telecare* 2018 Jun;24(5):356-364. [doi: [10.1177/1357633X17700552](https://doi.org/10.1177/1357633X17700552)] [Medline: [28463033](https://pubmed.ncbi.nlm.nih.gov/28463033/)]
19. 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](https://doi.org/10.2196/12340)] [Medline: [31293241](https://pubmed.ncbi.nlm.nih.gov/31293241/)]
20. von Storch K, Graaf E, Wunderlich M, Rietz C, Polidori MC, Woopen C. Telemedicine-assisted self-management program for type 2 diabetes patients. *Diabetes Technol Ther* 2019 Sep;21(9):514-521. [doi: [10.1089/dia.2019.0056](https://doi.org/10.1089/dia.2019.0056)] [Medline: [31287736](https://pubmed.ncbi.nlm.nih.gov/31287736/)]
21. 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](https://doi.org/10.2196/jmir.6688)] [Medline: [28526671](https://pubmed.ncbi.nlm.nih.gov/28526671/)]
22. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. *Cochrane Handbook for Systematic Reviews of Interventions Version 6.2*. Hoboken, NJ: John Wiley & Sons; Feb 2021.
23. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: [10.1136/bmj.315.7109.629](https://doi.org/10.1136/bmj.315.7109.629)] [Medline: [9310563](https://pubmed.ncbi.nlm.nih.gov/9310563/)]
24. Chau JP, Lee DT, Yu DS, Chow AY, Yu W, Chair S, et al. A feasibility study to investigate the acceptability and potential effectiveness of a telecare service for older people with chronic obstructive pulmonary disease. *Int J Med Inform* 2012 Oct;81(10):674-682. [doi: [10.1016/j.ijmedinf.2012.06.003](https://doi.org/10.1016/j.ijmedinf.2012.06.003)] [Medline: [22789911](https://pubmed.ncbi.nlm.nih.gov/22789911/)]
25. Chow SKY, Wong FKY. A randomized controlled trial of a nurse-led case management programme for hospital-discharged older adults with co-morbidities. *J Adv Nurs* 2014 Oct;70(10):2257-2271 [FREE Full text] [doi: [10.1111/jan.12375](https://doi.org/10.1111/jan.12375)] [Medline: [24617755](https://pubmed.ncbi.nlm.nih.gov/24617755/)]
26. De San Miguel K, Smith J, Lewin G. Telehealth remote monitoring for community-dwelling older adults with chronic obstructive pulmonary disease. *Telemed J E Health* 2013 Sep;19(9):652-657. [doi: [10.1089/tmj.2012.0244](https://doi.org/10.1089/tmj.2012.0244)] [Medline: [23808885](https://pubmed.ncbi.nlm.nih.gov/23808885/)]
27. Finkelstein SM, Speedie SM, Zhou X, Pothhoff S, Ratner ER. Perception, satisfaction and utilization of the VALUE home telehealth service. *J Telemed Telecare* 2011;17(6):288-292. [doi: [10.1258/jtt.2011.100712](https://doi.org/10.1258/jtt.2011.100712)] [Medline: [21844178](https://pubmed.ncbi.nlm.nih.gov/21844178/)]
28. Finlayson K, Chang AM, Courtney MD, Edwards HE, Parker AW, Hamilton K, et al. Transitional care interventions reduce unplanned hospital readmissions in high-risk older adults. *BMC Health Serv Res* 2018 Dec 12;18(1):956 [FREE Full text] [doi: [10.1186/s12913-018-3771-9](https://doi.org/10.1186/s12913-018-3771-9)] [Medline: [30541530](https://pubmed.ncbi.nlm.nih.gov/30541530/)]
29. Gellis ZD, Kenaley BL, Ten Have T. Integrated telehealth care for chronic illness and depression in geriatric home care patients: the Integrated Telehealth Education and Activation of Mood (I-TEAM) study. *J Am Geriatr Soc* 2014 May;62(5):889-895. [doi: [10.1111/jgs.12776](https://doi.org/10.1111/jgs.12776)] [Medline: [24655228](https://pubmed.ncbi.nlm.nih.gov/24655228/)]
30. Jolly K, Sidhu MS, Hewitt CA, Coventry PA, Daley A, Jordan R, et al. Self management of patients with mild COPD in primary care: randomised controlled trial. *BMJ* 2018 Jun 13;361:k2241 [FREE Full text] [doi: [10.1136/bmj.k2241](https://doi.org/10.1136/bmj.k2241)] [Medline: [29899047](https://pubmed.ncbi.nlm.nih.gov/29899047/)]
31. Kazawa K, Osaki K, Rahman MM, Moriyama M. Evaluating the effectiveness and feasibility of nurse-led distant and face-to-face interviews programs for promoting behavioral change and disease management in patients with diabetic nephropathy: a triangulation approach. *BMC Nurs* 2020;19:16 [FREE Full text] [doi: [10.1186/s12912-020-0409-0](https://doi.org/10.1186/s12912-020-0409-0)] [Medline: [32189998](https://pubmed.ncbi.nlm.nih.gov/32189998/)]
32. Kleinpell RM. Randomized trial of a discharge planning and telehealth intervention for patients aged 65 and older after coronary artery bypass surgery. *Int J Clin Cardiol* 2015 Aug 31;2(4):1-6. [doi: [10.23937/2378-2951/1410044](https://doi.org/10.23937/2378-2951/1410044)]
33. Oksman E, Linna M, Hörhammer I, Lammintakanen J, Talja M. Cost-effectiveness analysis for a tele-based health coaching program for chronic disease in primary care. *BMC Health Serv Res* 2017 Feb 15;17(1):138 [FREE Full text] [doi: [10.1186/s12913-017-2088-4](https://doi.org/10.1186/s12913-017-2088-4)] [Medline: [28202032](https://pubmed.ncbi.nlm.nih.gov/28202032/)]
34. Pecina JL, Hanson GJ, Van Houten H, Takahashi PY. Impact of telemonitoring on older adults health-related quality of life: the Tele-ERA study. *Qual Life Res* 2013 Nov;22(9):2315-2321. [doi: [10.1007/s11136-013-0361-5](https://doi.org/10.1007/s11136-013-0361-5)] [Medline: [23408299](https://pubmed.ncbi.nlm.nih.gov/23408299/)]
35. 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](https://doi.org/10.1001/archinternmed.2012.256)] [Medline: [22507696](https://pubmed.ncbi.nlm.nih.gov/22507696/)]
36. Wong FKY, Chow SKY, Chan TMF, Tam SKF. Comparison of effects between home visits with telephone calls and telephone calls only for transitional discharge support: a randomised controlled trial. *Age Ageing* 2014 Jan;43(1):91-97 [FREE Full text] [doi: [10.1093/ageing/aft123](https://doi.org/10.1093/ageing/aft123)] [Medline: [23978408](https://pubmed.ncbi.nlm.nih.gov/23978408/)]

37. Zobair KM, Sanzogni L, Sandhu K. Expectations of telemedicine health service adoption in rural Bangladesh. *Soc Sci Med* 2019 Oct;238:112485. [doi: [10.1016/j.socscimed.2019.112485](https://doi.org/10.1016/j.socscimed.2019.112485)] [Medline: [31476664](https://pubmed.ncbi.nlm.nih.gov/31476664/)]
38. Joo JY, Liu MF. Effectiveness of nurse-led case management in cancer care: systematic review. *Clin Nurs Res* 2019 Nov;28(8):968-991. [doi: [10.1177/1054773818773285](https://doi.org/10.1177/1054773818773285)] [Medline: [29726271](https://pubmed.ncbi.nlm.nih.gov/29726271/)]
39. Niu X, Chi J, Guo J, Ruan H, Zhang J, Tao H, et al. Effects of nurse-led web-based interventions on people with type 2 diabetes mellitus: A systematic review and meta-analysis. *J Telemed Telecare* 2021 Jun;27(5):269-279. [doi: [10.1177/1357633X211010019](https://doi.org/10.1177/1357633X211010019)] [Medline: [33966522](https://pubmed.ncbi.nlm.nih.gov/33966522/)]
40. Dunn T, Bliss J, Rylie I. The impact of community nurse-led interventions on the need for hospital use among older adults: An integrative review. *Int J Older People Nurs* 2021 Mar;16(2):e12361. [doi: [10.1111/opn.12361](https://doi.org/10.1111/opn.12361)] [Medline: [33486899](https://pubmed.ncbi.nlm.nih.gov/33486899/)]
41. Auger KA, Shah SS, Tubbs-Cooley HL, Sucharew HJ, Gold JM, Wade-Murphy S, Hospital-to-Home Outcomes Trial Study Group. Effects of a 1-time nurse-led telephone call after pediatric discharge: The H2O II randomized clinical trial. *JAMA Pediatr* 2018 Sep 01;172(9):e181482 [FREE Full text] [doi: [10.1001/jamapediatrics.2018.1482](https://doi.org/10.1001/jamapediatrics.2018.1482)] [Medline: [30039161](https://pubmed.ncbi.nlm.nih.gov/30039161/)]
42. Awoke MS, Baptiste D, Davidson P, Roberts A, Dennison-Himmelfarb C. A quasi-experimental study examining a nurse-led education program to improve knowledge, self-care, and reduce readmission for individuals with heart failure. *Contemp Nurse* 2019 Feb;55(1):15-26. [doi: [10.1080/10376178.2019.1568198](https://doi.org/10.1080/10376178.2019.1568198)] [Medline: [30632457](https://pubmed.ncbi.nlm.nih.gov/30632457/)]
43. Harrison MB, Browne GB, Roberts J, Tugwell P, Gafni A, Graham ID. Quality of life of individuals with heart failure: a randomized trial of the effectiveness of two models of hospital-to-home transition. *Med Care* 2002 Apr;40(4):271-282. [Medline: [12021683](https://pubmed.ncbi.nlm.nih.gov/12021683/)]
44. Krumholz HM, Amatruda J, Smith GL, Mattera JA, Roumanis SA, Radford MJ, et al. Randomized trial of an education and support intervention to prevent readmission of patients with heart failure. *J Am Coll Cardiol* 2002 Jan 02;39(1):83-89 [FREE Full text] [doi: [10.1016/s0735-1097\(01\)01699-0](https://doi.org/10.1016/s0735-1097(01)01699-0)] [Medline: [11755291](https://pubmed.ncbi.nlm.nih.gov/11755291/)]

Abbreviations

OR: odds ratio

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

QoL: quality of life

RCT: randomized controlled trial

SMD: standardized mean difference

Edited by A Mavragani; submitted 09.07.21; peer-reviewed by B Seah, M Stein, E Vidoni; comments to author 28.08.21; revised version received 18.10.21; accepted 18.01.22; published 21.03.22.

Please cite as:

Wong AKC, Bayuo J, Wong FKY, Yuen WS, Lee AYL, Chang PK, Lai JTC

Effects of a Nurse-Led Telehealth Self-care Promotion Program on the Quality of Life of Community-Dwelling Older Adults: Systematic Review and Meta-analysis

J Med Internet Res 2022;24(3):e31912

URL: <https://www.jmir.org/2022/3/e31912>

doi: [10.2196/31912](https://doi.org/10.2196/31912)

PMID: [35311680](https://pubmed.ncbi.nlm.nih.gov/35311680/)

©Arkers Kwan Ching Wong, Jonathan Bayuo, Frances Kam Yuet Wong, Wing Shan Yuen, Athena Yin Lam Lee, Pui King Chang, Jojo Tsz Chui Lai. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 21.03.2022. 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.

Review

Inequities in Health Care Services Caused by the Adoption of Digital Health Technologies: Scoping Review

Rui Yao^{1*}, MS; Wenli Zhang^{1*}, MS; Richard Evans², PhD; Guang Cao¹, MS; Tianqi Rui¹, MS; Lining Shen^{1,3,4}, PhD

¹School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science & Technology, Wuhan, China

²Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada

³Hubei Provincial Research Center for Health Technology Assessment, Wuhan, China

⁴Institute of Smart Health, Huazhong University of Science & Technology, Wuhan, China

*these authors contributed equally

Corresponding Author:

Lining Shen, PhD

School of Medicine and Health Management

Tongji Medical College

Huazhong University of Science & Technology

No 13 Hangkong Road

Wuhan, 430030

China

Phone: 86 (027)83692730

Email: sln2008@hust.edu.cn

Abstract

Background: Digital health technologies (ie, the integration of digital technology and health information) aim to increase the efficiency of health care delivery; they are rapidly adapting to health care contexts to provide improved medical services for citizens. However, contrary to expectations, their rapid adoption appears to have led to health inequities, with differences in health conditions or inequality in the distribution of health care resources among different populations.

Objective: This scoping review aims to identify and describe the inequities of health care services brought about by the adoption of digital health technologies. The factors influencing such inequities, as well as the corresponding countermeasures to ensure health equity among different groups of citizens, were also studied.

Methods: Primary studies and literature, including articles and reviews, published in English between 1990 and 2020 were retrieved using appropriate search strategies across the following three electronic databases: Clarivate Analytics' Web of Science, PubMed, and Scopus. Data management was performed by two authors (RY and WZ) using Thomson Endnote (Clarivate Analytics, Inc), by systematically screening and identifying eligible articles for this study. Any conflicts of opinion were resolved through discussions with the corresponding author. A qualitative descriptive synthesis was performed to determine the outcomes of this scoping review.

Results: A total of 2325 studies were collected during the search process, of which 41 (1.76%) papers were identified for further analysis. The quantity of literature increased until 2016, with a peak in 2020. The United States, the United Kingdom, and Norway ranked among the top 3 countries for publication output. Health inequities caused by the adoption of digital health technologies in health care services can be reflected in the following two dimensions: the inability of citizens to obtain and adopt technology and the different disease outcomes found among citizens under technical intervention measures. The factors that influenced inequities included age, race, region, economy, and education level, together with health conditions and eHealth literacy. Finally, action can be taken to alleviate inequities in the future by government agencies and medical institutions (eg, establishing national health insurance), digital health technology providers (eg, designing high-quality tools), and health care service recipients (eg, developing skills to access digital technologies).

Conclusions: The application of digital health technologies in health care services has caused inequities to some extent. However, existing research has certain limitations. The findings provide a comprehensive starting point for future research, allowing for further investigation into how digital health technologies may influence the unequal distribution of health care services. The interaction between individual subjective factors as well as social support and influencing factors should be included in future studies. Specifically, access to and availability of digital health technologies for socially disadvantaged groups should be of paramount importance.

KEYWORDS

health inequities; digital health technologies; health care services; socially disadvantaged groups; scoping review; mobile phone

Introduction

Background

An evolution in health care services is occurring across the globe in response to an explosion in readily available digital technologies. The adoption of digital technologies as a means for citizens to access health and social care is accelerating at an unprecedented pace, pushing patient-centered care toward digital health [1]. Many countries and organizations are paying greater attention to digital health, which has seen a sharp increase in the release of health policies and reports, such as the United Kingdom's *Digital Strategy* published in 2012 [2], the European Union's *Europe's Digital Decade* and *Digital Europe Programme* [3], the World Health Organization's *Draft Global Strategy for Digital Health (2020-2025)* [4], China's *Outline of Healthy China 2030 Plan* [5], and so on. Digital health refers to the use of readily available information and communication technologies for the following: to provide patients with preventive services, treatment, and education; to promote disease tracking and monitoring; and to enable consumers to participate in health care services [6,7]. Digital health is the integration of digital technology and health information with the aim of increasing the efficiency of health care delivery and improving the health of patients [8]. Digital health technology is the adoption of digital technology in health, with examples being seen in electronic health records, telemedicine or telehealth services, robotics, and eHealth, along with mobile health supported by the use of smartphones, wearables, mobile apps, and various monitoring devices [9-11].

Lupton [12] had said that "Digital health technologies are positioned to enable people to effectively become 'managers' of their own health and healthcare." In our internet-enabled world, the use of digital health technologies is becoming the core of health care delivery. Studies have shown that digital health technology can improve health literacy, enhance patient participation in health care, enable patients to better manage their own health, and improve health care efficiency, especially in patients with chronic diseases [13,14]. Digital health technology interventions, that is, those delivered through digital technologies, such as smartphones and websites [15], can improve health care delivery and contribute to the *triple aim* of health care, that is, better care, better health outcomes, and reduction in medical spending [16]. However, compared with expectations, the rapid development of digital health technologies has led to health inequities.

Health inequities refer to differences in health conditions or the distribution of health care resources among different populations because of social conditions, such as the citizens' place of birth, growth, life, or work [17]. Although digital health technologies are being adopted rapidly, it is likely that those who do not use the internet or mobile devices regularly or have difficulty in using them, such as older adults, those living in low-income regions, and people in remote areas with poor internet

connectivity, will be forgotten [18-20]. This phenomenon not only represents inequities among income, education, and age groups and between the healthiest and least healthy [17] but also represents inequities in access to and availability of technology, which is a continuing barrier to the use of digital health services [21].

The potential of technologies to induce health inequities has been widely recognized [22]. As early as 2016, the World Bank's Information Industry Report identified that information technology innovations have the potential to lead to new inequities [20]. The report stated that those who are wealthy and better educated are well positioned to take advantage of the internet. However, many global citizens do not have access to the internet. In some regions where women have low socioeconomic status (SES), they are discouraged from going on the web and do not have access to cell phones [20]. In addition, there are still people in parts of the world who are illiterate and do not benefit from access to the internet. Some studies found that mobile health interventions can exacerbate treatment disparities [23,24]. Digital health technology interventions work better for those who are already better off—a situation that can induce inequities. This phenomenon is well established in public health and is referred to as *Intervention-Generated Inequalities* [19]. Socially disadvantaged groups [25] have more challenges in access to and availability of digital health technologies [26], which may lead to more severe health inequities.

Objectives

From an ethical perspective, health equity is more important than health inequity as the latter can have negative social and economic consequences [27]. The databases of Cochrane, JBI Evidence Synthesis, and PROSPERO (the international prospective register of systematic reviews administered by the University of York's Centre for Reviews and Dissemination) show that there is no literature review to date on inequities caused by the application of digital health technologies in health care services. The aim of this scoping review is to systematically review and synthesize information on health inequities in health care delivery resulting from digital health technologies and to provide insights for future research and practice. Such a review can provide a better understanding of the health inequities caused by digital health technologies, the influencing factors, and countermeasures and can inform future corresponding policy decision-making to ensure health equities among different citizen groups, and thereby achieve social equity and stability.

Methods

Overview

This scoping review used the framework of Arksey and O'Malley [28], which comprises the following five stages: identifying the research question; identifying relevant studies; study selection; data extraction and analysis; and collating,

summarizing, and reporting the results. The reporting of the scoping review was guided by the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist; for details of items included in the checklist, please see [Multimedia Appendix 1](#) [29]. In addition, the study was registered in an open science framework and an independent registry. The registration type of this scoping review is open-ended and the registration digital object identifier is 10.17605/OSF.IO/A5R7F.

Search Strategy

The electronic databases Clarivate Analytics' Web of Science, PubMed, and Scopus were searched for articles published in English between 1990 (in the late 1990s, the combination of

medical care and technology gave birth to a new field called *eHealth* [30]) and 2020. Two coauthors (RY and WZ) developed and performed a Boolean search strategy based on SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, Research) type—a search tool used to identify relevant qualitative and mixed method studies [31]. The search frame included phenomenon of interest (digital technologies, such as *ehealth* or *telehealth**), evaluation (*inequit**), and research type (primary studies and literature including articles or reviews; see [Table 1](#) for details). Through a review of the keywords that appeared in the title or summary field of each article, a list of literature was retrieved. The title information was then exported to EndNote X9 (Clarivate Analytics, Inc) for evaluation to promote the selection process and collaboration among reviewers.

Table 1. Database search strategy.

Database	Search strategy	Number of results
Web of science	<i>TS= (eHealth OR telehealth* OR telemedicine OR mHealth OR mobile health OR health IT OR health information technolog* OR health informat* OR digital health* OR digital health technolog* OR [ICT AND (health* OR healthcare settings OR healthcare delivery OR healthcare service*)] OR [technolog* AND (health* OR healthcare settings OR healthcare delivery OR healthcare service*)]) AND TS= inequit*. Time span=January 1, 1990, to December 31, 2020.</i>	910
PubMed	<i>(1990[Date—Publication]: 2020[Date—Publication]) AND (eHealth[Title/Abstract] OR telehealth*[Title/Abstract] OR telemedicine[Title/Abstract] OR mHealth[Title/Abstract] OR mobile health[Title/Abstract] OR health IT[Title/Abstract] OR health information technolog*[Title/Abstract] OR health informat*[Title/Abstract] OR digital health*[Title/Abstract] OR digital health technolog*[Title/Abstract] OR [ICT(Title/Abstract) AND (health*[Title/Abstract] OR health care settings[Title/Abstract] OR health care delivery[Title/Abstract] OR health care service*[Title/Abstract])]) OR [technolog*(Title/Abstract) AND (health*[Title/Abstract] OR health care settings[Title/Abstract] OR health care delivery[Title/Abstract] OR health care service*[Title/Abstract])]) OR [telemedicine(MeSH Terms)] AND (inequit*[Title/Abstract]).</i>	566
Scopus	<i>(TITLE-ABS-KEY[eHealth] OR TITLE-ABS-KEY[telehealth*] OR TITLE-ABS-KEY[telemedicine] OR TITLE-ABS-KEY[mHealth] OR TITLE-ABS-KEY[mobile health] OR TITLE-ABS-KEY[health IT] OR TITLE-ABS-KEY[health information technolog*] OR TITLE-ABS-KEY[health informat*] OR TITLE-ABS-KEY[digital health*] OR TITLE-ABS-KEY[digital health technolog*] OR [TITLE-ABS-KEY(ICT) AND (TITLE-ABS-KEY[health*] OR TITLE-ABS-KEY[healthcare settings] OR TITLE-ABS-KEY[healthcare delivery] OR TITLE-ABS-KEY[healthcare service*])] OR [TITLE-ABS-KEY(technolog*) AND (TITLE-ABS-KEY[health*] OR TITLE-ABS-KEY[healthcare settings] OR TITLE-ABS-KEY[healthcare delivery] OR TITLE-ABS-KEY[healthcare service*])] AND TITLE-ABS-KEY(inequit*) AND PUBYEAR AFT 1989 AND PUBYEAR BEF 2021.</i>	849

Inclusion Criteria

Studies were considered eligible if they met the following criteria: (1) published in English between the years 1990 and 2020, (2) either a primary study or literature review, and (3) discussed health inequities related to digital health technology interventions or explored the influencing factors for digital health inducing health inequities or the countermeasures to alleviate health inequities.

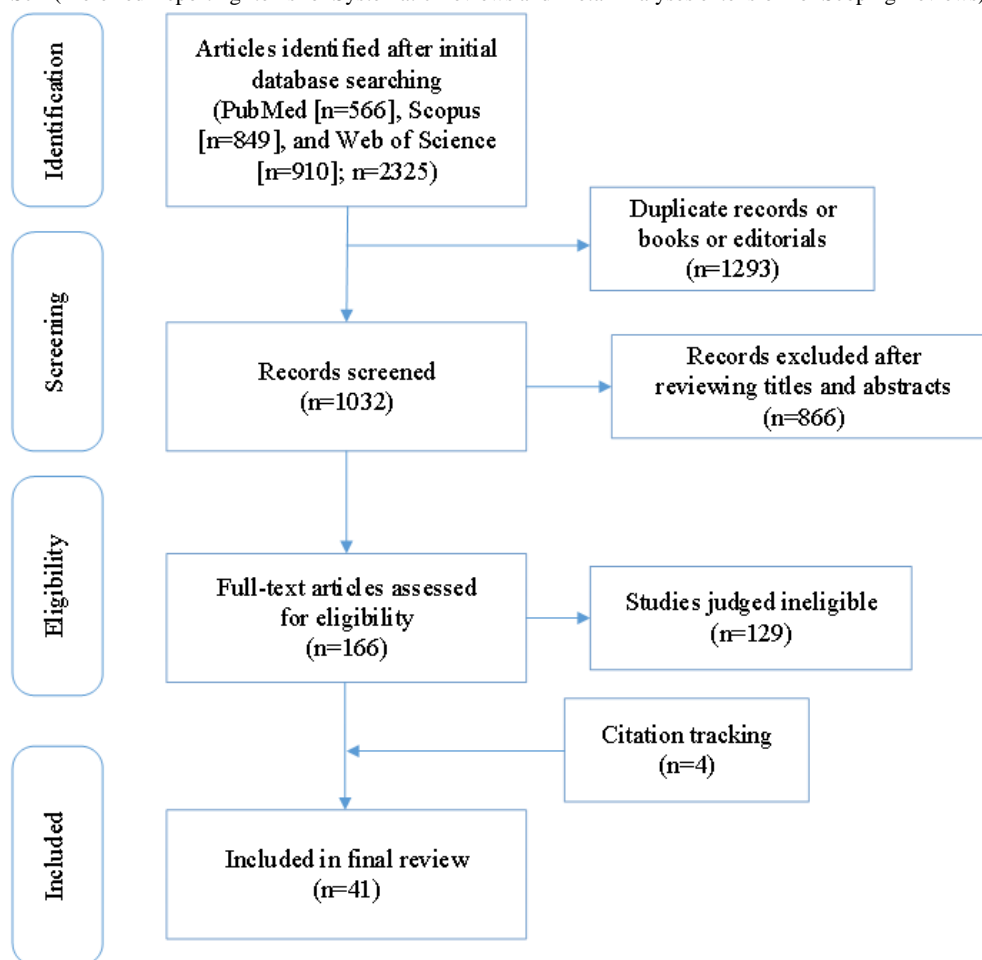
Exclusion Criteria

Studies considered ineligible included those that only related to the following: (1) books or book sections or editorials, commentary, and columns; (2) studies beyond the reach of the full text; (3) public health intervention and policy intervention measures; (4) the design of health technology, service systems, or frameworks to make up for health differences; or (5) studies that did not explore the relationship between digital health technologies and health inequities.

Study Selection

After designing the search strategy, inclusion criteria, and exclusion criteria, the literature was reviewed according to the PRISMA-ScR process [29]. A total of 2325 records were collected from the following electronic databases: Clarivate Analytics' Web of Science, PubMed, and Scopus. To review the relevance of the literature, coauthors RY and WZ screened the titles and abstracts of all remaining records after the removal of duplicates. All full texts were read and analyzed by 2 individual researchers (RY and WZ), and individual data extraction forms were then merged into a single, unifying document that was used for the interpretation and presentation of results. Discrepancies were adjudicated by the corresponding authors. Next, the full article text of the retrieved results was examined by RY and WZ based on the inclusion and exclusion criteria. Of the 2325 papers, 41 (1.76%) papers were identified for the systematic analysis, as illustrated in [Figure 1](#).

Figure 1. PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) procedural flowchart.



Data Extraction and Analysis

A bespoke data extraction form was used to systematically extract information relevant to the descriptive information. Data were extracted from each article to describe the following: (1) the characteristics of each study (including author, year of publication, publication country, and methods used), (2) the concrete embodiment of the inequity of health care services brought about by the application of digital health technology, (3) the factors contributing to the inequities in health care services resulting from the adoption of digital health technology, and (4) the measures used to mitigate health inequities.

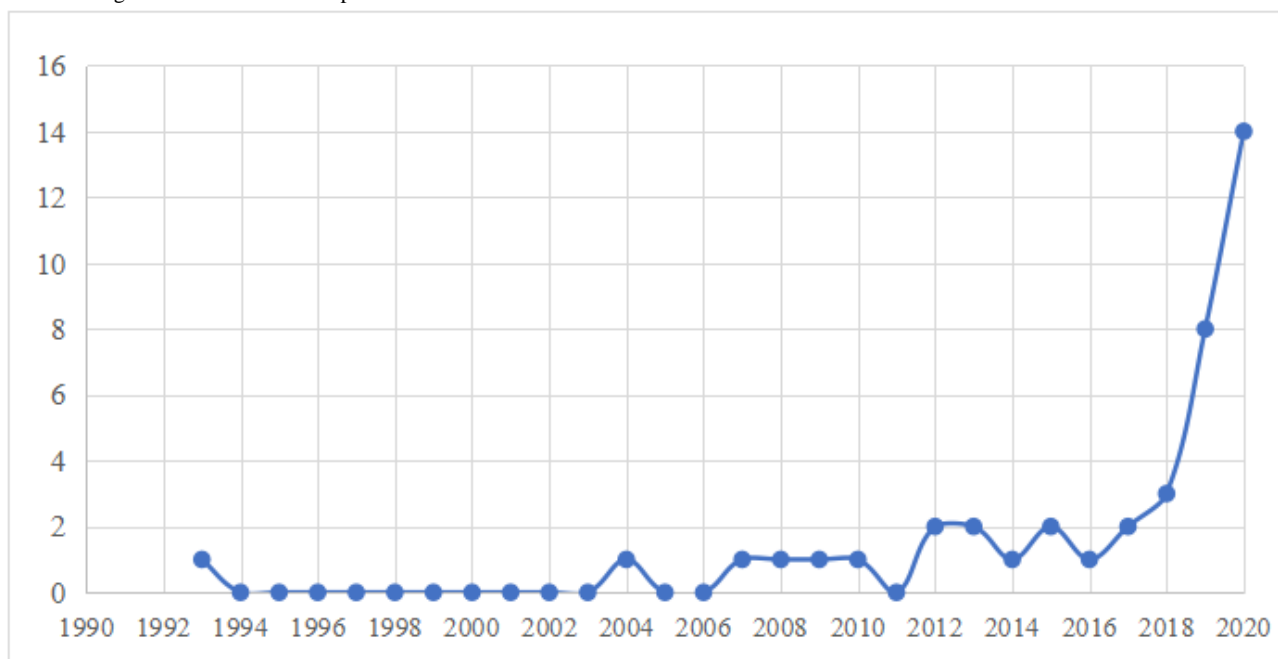
Furthermore, a qualitative descriptive synthesis [30] was performed to determine the outcomes of this scoping review.

Results

Overview of the Included Studies

Of the 41 papers retained, the earliest publication related to health inequities caused by digital health technologies was published in 1993. Thereafter, the number of relevant publications remained at 0 from 1994-2003, fluctuated between 0 and 2 from 2004-2016, and grew after 2016, reaching a peak of 14 in 2020, as shown in Figure 2.

Figure 2. Change in trend of number of publications over time.



The United States published the greatest number of papers on the topic, accounting for 46% (19/41), followed by the United Kingdom (8/41, 20%), and Norway (3/41, 7%), whereas other countries published only 1 or 2 papers, as shown in [Table 2](#). Of the 41 records, 6 (15%) were literature reviews, 17 (41%)

records used a quantitative approach, 15 (37%) records used a qualitative approach, and 3 (7%) records used mixed methods. Specific information on the included studies is presented in [Table 3](#).

Table 2. Statistics of literature publications in different countries (N=41).

Country	Number of literatures, n (%)
Australia	1 (2)
Bangladesh	1 (2)
Canada	2 (5)
Indonesia	1 (2)
Israel	1 (2)
Italy	2 (5)
Korea	1 (2)
Netherlands	1 (2)
Norway	3 (7)
Switzerland	1 (2)
United Kingdom	8 (20)
United States	19 (46)

Table 3. Overview of included studies.

Authors	Year	Country	Research method	Manifestations of health inequities	Influencing factors of health inequities	Countermeasures for health inequities
Yang [32]	1993	Korea	Quantitative approach	Access to health care resource: medical insurance access	Income	Countermeasures of government agencies: establish national health insurance
Steiger et al [33]	2004	Switzerland	Qualitative approach	N/A ^a	Age and SES ^b	Countermeasures of technology providers: design high-quality websites
Viswanath et al [34]	2007	United States	Qualitative approach	Access to health care resource: technical access and availability	SES	Countermeasures of government agencies: structural adjustment at the policy system level
Kim et al [35]	2009	United States	Quantitative approach	Access to health care resource: technology acquisition	Economy, age, and eHealth literacy	Countermeasures of government agencies: free PCs, internet connections, and help from nursing students and accommodation staff
Andreassen et al [36]	2010	Norway	Qualitative approach	Health and disease outcome: disease mortality	SES and health literacy	N/A
Goldberg [37]	2012	United States	Qualitative approach	Health and disease outcome: disease risk	SES	Countermeasures of government agencies: give priority to people with diseases
Jones [38]	2013	United Kingdom	Quantitative approach	N/A	Health condition and economic barriers	Countermeasures of government agencies and health care service receivers: provide technical support, medical institutions take actions or volunteer services
Jennings et al [23]	2013	United States	Overview	Access to health care resource: technology acquisition	N/A	N/A
McAuley [18]	2014	United Kingdom	Combination of qualitative and quantitative	Access to health care resource: technology acquisition; health and disease outcome: increased risk of disease	SES and age	Countermeasures of health care service receivers: develop information and skills for access to digital technology
Albright et al [39]	2015	United States	Quantitative approach	Access to health care resource: technology acquisition	Age	N/A
Matteucci [40]	2015	Italy	Qualitative approach	Health and disease outcome: disease risk	Health condition	Countermeasures of government agencies: role of community and health professionals
Mierlo et al [41]	2016	United Kingdom	Quantitative approach	Access to health care resource: digital participation	N/A	N/A
Latulippe et al [42]	2017	Canada	Overview	Health and disease outcome: disease prevalence, mortality; access to health care resource: technology acquisition	eHealth literacy, age, SES, rural, and sexual orientation	Countermeasures of technology providers: universal eHealth care tool and technology design

Authors	Year	Country	Research method	Manifestations of health inequities	Influencing factors of health inequities	Countermeasures for health inequities
Hosseinpoor et al [43]	2017	Indonesia	Qualitative approach	Health and disease outcome: Health outcomes; access to health care resource: technical services	N/A	Countermeasures of government agencies: health inequities monitoring embedded in national health information system
Veinot et al [19]	2018	United States	Qualitative approach	N/A	Education level and health literacy	Countermeasures of government agencies and technology providers: invest in resources and time or technical design that is easy to understand and health professionals publish relevant policies
Bol et al [44]	2018	Netherlands	Quantitative approach	Access to health care resource: digital health technologies	Age, education, and eHealth literacy	Countermeasures of government agencies: develop more detailed strategies to bridge the digital divide
Weiss et al [45]	2018	Italy	Overview	Health and disease outcome: disease morbidity and mortality	SES, health condition, age, and race	N/A
Gann [46]	2019	United Kingdom	Qualitative approach	Access to health care resource: technical services	Age, SES, and rural	Countermeasures of health care service receivers: the role of libraries in providing resources
Toscos et al [47]	2019	United States	Quantitative approach	Access to health care resource: digital health technologies	Age and SES	Countermeasures of government agencies and medical institutions: health science researchers should consider the population with needs and insufficient coverage
Sherman et al [1]	2019	United States	Qualitative approach	Health and disease outcome: health outcomes	Race	N/A
Parker et al [48]	2019	United States	Quantitative approach	Health and disease outcome: health outcomes	N/A	N/A
Hansen et al [49]	2019	Norway	Quantitative approach	Access to health care resource: technical services; health and disease outcome: disease morbidity and mortality	SES, age, and health condition	N/A
Baum et al [50]	2012	Australia	Qualitative approach	Access to health care resource: technical services	SES and health literacy	Countermeasures of government agencies: health promoters must understand and explain the complex interaction among digital literacy, health literacy, and basic literacy

Authors	Year	Country	Research method	Manifestations of health inequities	Influencing factors of health inequities	Countermeasures for health inequities
Banerjee [51]	2019	United Kingdom	Qualitative approach	N/A	Race	Countermeasures of government agencies and technology providers: health professionals should receive better training in the evidence and use of DHI ^c ; the participant should be consulted during the design and implementation phase of the technology
Rich et al [52]	2019	United Kingdom	Qualitative approach	N/A	Economy	Policy documents should focus on inequities
Ahmed et al [53]	2020	Bangladesh	Combination of qualitative and quantitative	Access to health care resource: technical services	Age, SES, and digital health awareness and skills	Countermeasures of health care service receivers: raising public awareness and political incentives
Glied et al [54]	2008	United States	Quantitative approach	Health and disease outcome: disease mortality	Education	N/A
Fujioka et al [55]	2020	Canada	Overview	Health and disease outcome: health outcomes	Race, income, and health condition	N/A
Khilnani et al [56]	2020	United States	Qualitative approach	Health and disease outcome: health outcomes	Age, SES, rural, and digital resources	N/A
DeGuzman et al [57]	2020	United States	Quantitative approach	Access to health care resource: technical services	Rural and broadband coverage	Countermeasures of health care service receivers: public libraries provide equipment and hardware support
Gann [58]	2020	United Kingdom	Qualitative approach	Health and disease outcome: health outcomes	Economy and age	Countermeasures of government agencies: encourage enterprises to donate tablets, smartphones, and laptops, and provided tablet devices to hospitals, nursing homes, and hospice care institutions
Bommakanti et al [59]	2020	United States	Quantitative approach	N/A	Age and SES	N/A
Weiss et al [60]	2020	Norway	Quantitative approach	Health and disease outcome: health outcomes	SES	N/A
Nittas et al [24]	2020	United Kingdom	Overview	Access to health care resource: access and availability of technology	Race, SES, and place of residence	N/A
Karri et al [61]	2020	United States	Quantitative approach	Access to health care resource: technical services	SES, race, age, and health condition	N/A
Jaffe et al [62]	2020	Israel	Quantitative approach	N/A	Age and rural	N/A
DeGuzman et al [63]	2020	United States	Combination of qualitative and quantitative	Access to health care resource: digital health technologies	Digital literacy and skills and provision of broadband facilities	N/A

Authors	Year	Country	Research method	Manifestations of health inequities	Influencing factors of health inequities	Countermeasures for health inequities
Sun et al [64]	2020	United States	Qualitative approach	Access to health care resource: technical services	Digital literacy	Countermeasures of government agencies: strengthen the national digital health strategy and the governance of human rights-oriented digital health technologies at the national level
Ukoha et al [65]	2019	United States	Quantitative approach	Health and disease outcome: health outcomes and care	Socioeconomic factors, age, and race	N/A
Hamideh et al [66]	2020	United States	Review	N/A	Age, race, income, rural, health condition, and health literacy	N/A
Erhunmwunsee et al [67]	2020	United States	Quantitative approach	Health and disease outcome: treatment results	Income level, rural location of the hospital and insurance status	N/A

^aN/A: not available.

^bSES: socioeconomic status.

^cDHI: digital health intervention.

Inequities Caused by Digital Health Technologies

Health inequities were reflected in 2 aspects. The first was access to and availability of digital health technologies by different social groups, with 19 (46%) of the 41 studies describing unfair distribution. Among these 19 studies, 9 (47%) studies referred to a lack of network infrastructure (eg, internet broadband access, satellite towers, and power) and the availability of smartphones or computers when using digital health technologies [32,39,41,43,47,53,57,63,64]. A study pointed out that access to home-based telemedicine was inequitably distributed in the United States owing to the limited reach of fixed broadband in rural areas [57]. Of the 41 studies, 12 (29%) studies reported an unfair phenomenon of digital exclusion [19,23,24,34,35,41,44,46,49,50,61,64]. A study described how, without proper planning and safeguards, digital health technologies can contribute to expanding health inequity, widening the *digital divide* that separates those who can from those who cannot access such interventions [64].

The remaining (17/41, 41%) studies discussed the related health outcomes caused by the lack of or limited access to digital health technologies [1,18,36,37,40,42,43,45,48,49,54-56,58,60,65,67]. Researchers have reported different health outcomes following the introduction of digital health technology interventions, including disease incidence rates and mortality, with particular attention paid to chronic diseases and Black groups [49,56,58,60]. For example, the average blood sugar level of diabetic patients who use innovative health technologies generally drops [49], and the use of robotic lobectomy is limited by sociodemographic factors, leading to significant treatment differences in patients with lung cancer [67].

Factors Leading to Health Inequities

Most (37/41, 90%) papers reported the factors that influence health inequities brought about by digital health services. Of these 37 studies, 18 (49%) studies reported the important role

of age in determining whether patients use or do not use digital health technologies, especially among the older adults [18,33,35,39,42,44-47,49,53,56,58,59,61,62,65,66]. Race and ethnicity have long been considered as one of the *causes of the causes* of health inequities [51]; the acknowledged health benefits of eHealth or mobile health initiatives have seen limited application among Black communities [1,24,42,45,51,55,61,65,66]. In addition, some studies (8/41, 20%) reported that citizens living in rural areas were affected by poor access to and availability of digital health technologies resulting from limited internet broadband coverage [24,42,45,46,56,57,62,66]. In rural areas with multiple health barriers, the unequal distribution of health information is more complex and, therefore, may exacerbate rather than narrow the gap [68]. Furthermore, SES (eg, education and income) [69] was also discussed. Some (18/41, 44%) studies have reported differences in the acquisition and adoption of digital health technologies by different groups of citizens based on a combination of income and education [18,24,33,34,36,37,42,44,46,47,49,50,53,56,59-61,64], whereas others (13/41, 32%) have independently explored the 2 dimensions of income [32,35,38,52,55,58,65-67] and educational attainment [19,44,54,62].

In addition to the above factors, poor health conditions have been observed to hinder access to digital health technologies at a physical level [38,40,45,49,55,61] or a lack of confidence in health advice and health decision-making at a psychological level [38,42]. A literature review found that some people may have difficulty learning to use the internet because of visual or hearing impairment, arthritis, or hand movement difficulties [38]. Studies (11/41, 27%) have also pointed out that the so-called low eHealth literacy (defined as “the ability to seek, find, understand, and appraise health information from electronic sources and apply knowledge gained to addressing or solving a health problem” [70]), including a lack of awareness of digital health services, literacy, and skills, as well as the understanding

of the operation mode and availability of digital health, may reduce acceptance of digital health technologies [20,35,36,38,41,42,50,53,63,64,66].

Countermeasures Lessening Health Inequities

Measures of Government Agencies and Medical Institutions

Government agencies and medical institutions are required to alleviate the inequities in digital health technology interventions. This study identified that human rights-oriented digital health technologies should be strengthened at the national level [64] and new research trajectories for policy documents must explore how culture, practice, and power relationships influence access, availability, and engagement with digital health technologies [34,52]. Furthermore, those who develop and apply these interventions, including policymakers and public health professionals, should have a better understanding of whether, how, and under what circumstances digital health technology interventions can overcome inequities and realize their potential [19,40,51,61]. Studies (2/41, 5%) have also recommended that the monitoring of health inequities be embedded within national health information systems and suggested that hybrid web-based and offline interventions can ameliorate health inequities [20,43].

Studies (4/41, 10%) reported that government agencies and medical institutions should provide help with resources, whereby government agencies could establish universal health insurance [32], provide free PC and internet connection help [35,38], and provide limited resources to treat disabled populations, such as those with brain injuries [37]. At the same time, government agencies should ensure increased public funding input [65] and encourage companies to donate tablets, smartphones, and laptops to hospitals, care homes, and hospice facilities [58] to make the use of services more accessible to socially disadvantaged groups, such as families who were not originally connected to the network [47].

Actions of Digital Health Technology Providers

Digital health services are inseparable from the design and development of targeted digital health technology. A study in Switzerland [33] proposed the adoption of high-quality internet portals with hashtags to promote autonomous learning and use among citizens. A study pointed out that most digital health technologies were not designed for socially disadvantaged groups, such as older adults or those with limited health literacy skills [47]. Therefore, the individual needs of groups should be carefully considered, including those of groups that need to be consulted during the design and implementation stage [42,51]; resources and time should be invested in the design of easy-to-understand language [19] and bespoke digital health tools should be created and popularized to maximize acceptability [51].

Recommended Measures for Health Care Service Receivers

The eHealth literacy of health care service receivers is an important factor hindering access to digital health services. Addressing health inequities depends on increasing public

awareness [53]. Some studies (5/41, 12%) suggested that relevant technical tools and volunteers can be provided [1,38] to develop patients' confidence and skills in digital technology access and improve their eHealth literacy [18,38,50,58]. In addition, some studies (2/41, 5%) reported on the important role of libraries. As a special public space, libraries have unique benefits, including auxiliary digital access, health information resources and services [46,57], and the voluntary services they provide, which can help patients improve their health literacy [46].

Discussion

Principal Findings

This scoping review identifies and describes the health inequities in health care services brought about by the adoption of digital health technologies. The evolution of publications over time and their distribution by country in the included literature reflects the concerns of related researchers to some extent. The rate at which the related literature has increased over time is consistent with the developing trend in health information technology. The included literature that has been published from 12 countries indicate the specific prevalence of health inequities on a global scale. Among them, the United States and the United Kingdom have published the most papers, mainly because they were the first to invest heavily in digital health information. Accordingly, health inequities caused by digital health technologies have been quickly noticed and studied.

The health inequities caused by digital health technologies highlight that not all citizens have equal access to and use of interventions, resulting in different health outcomes for the population. From our analysis of the studies investigated, we identified that these indicators, including age, race, SES, health conditions, eHealth literacy, and geographic location, all affect health inequity among groups. First, most theoretical frameworks that are widely used to understand the adoption of new technologies by users, such as *the technology acceptance model*, *the extended unified theory of acceptance and use of technology*, and the diffusion of innovations theory, have demonstrated potential individual differences by considering factors such as age and race. Regarding *age*, we found that older patients had less ownership, which was consistent with other studies [59,71]. During the COVID-19 pandemic, mobile smart devices have been the most effective way for people to seek medical treatment. The older adults may be unable to use mobile devices to register health codes or even for medical service preregistration. However, their long-term demand for medicines is higher [46], reflecting the gap in the use of digital health technologies among different age groups. Rogers [72], in his book *Diffusion of Innovations*, proposed that the older adults belong to laggards; they are relatively conservative when they encounter technological innovation and they feel skeptical and cautious about new things, which may affect their health. Meanwhile, in terms of *gender*, previous studies have reported a general trend that women are more involved in health issues, eHealth, and social media [49]; however, there are also studies that report that women in Africa are the least likely to use digital health technologies worldwide [46]. Moreover, previous studies

have found that gender and some ethnic differences in internet use may have disappeared among the general population [73,74] but differences will still expand with age. Health inequities have gradually become a social issue that has received attention from government decision-making departments and related institutions. The National Institutes of Health added a health information section to its official website, which includes links to many health care websites, such as child and adolescent health, men's health, women's health, minority health, and older adult health, to provide users and inquirers with the ability to obtain web-based health information.

Second, health inequities caused by *socioeconomic status* are prevalent throughout society, and digital health technologies are accessed to a greater degree by individuals with higher social standing. Link [75] referred to SES as a fundamental factor in health inequities. Social causality provides a theoretical explanation for health inequities and maintains that SES, such as education and income, are an important cause of health inequities. The *knowledge gap* hypothesis [76] holds that as the mass media increasingly disseminate information to society; higher SES groups access information at a faster rate than lower SES groups, so the knowledge gap between these groups will expand. On the basis of existing studies, the main reasons for health inequities caused by SES are as follows: (1) increased income can improve living conditions and access to digital health technologies and (2) a good educational level provides advantages in the adoption of digital health technologies [77]. As early as 20 years ago, Gordon Brown, the United Kingdom's Chancellor of the Exchequer, announced a plan to *shrink the digital divide*, proposing the provision of free internet access for poor communities, free information technology training, and 100,000 second-hand computers for low-income families. As early as 2010, the United States enacted the *Affordable Care Act*, which promoted access to and availability of health care information resource services by inquirers, especially those that were uninsured and low-income [78].

Third, as health literacy in the digital era may be considered a prerequisite for solving health problems in web-based environments [44], it requires people to better understand mobile health devices such as mobile health apps and how to obtain health information through the internet. In addition, the level of health literacy in the digital era determines the severity of health inequities to a large extent and the level of health literacy in the digital era is not determined by individuals alone. Compared with individual-level behavioral interventions, policy- and system-level interventions often have a greater impact on a population's health literacy in the digital era [34,79]. In February 2000, the US government released the report *From the Digital Divide to Digital Opportunity*, pointing out that the importance of bridging the digital divide lies in the popularization of professional technology, skills training, and the practicality of network content.

As a result, it can be found that it is necessary for stakeholders to take some measures to jointly build patient-centric digital health technology services to reduce inequities, promoting greater patient access to and use of high-quality digital health technology services. First, *policymakers and medical decision-makers* should promote digital health technologies to

poverty-stricken areas and provide low-income patients with free and high-quality health information resources and services, as appropriate [56]. Second, *medical institutions* should simplify the web-based service process and create channels for family members, relatives, friends, and physicians to make appointments. Manual service windows, such as registration, payment, and printing test results, should be reserved to deliver a combination of web-based and offline medical services [24]. Third, *related institutions* should increase investment in education and train individuals in health literacy or health information literacy to improve health outcomes and reduce health inequities [44]. Meanwhile, it is suggested that *relevant departments* should promote and encourage the older adults, the poor, the poorly educated, and other socially disadvantaged groups to use the internet and enrich and standardize the health information and knowledge of the internet to improve their eHealth literacy [18,38]. In view of the difficulties in the application of the internet in daily life, industry training institutions and experts should be organized to carry out special training to improve their operational ability with digital health technologies. Finally, when designing tools, *health care service providers* must participate in the development of assistive technologies, simplifying operations to help individuals who are older, have poor electronic literacy, or are poorly educated to overcome difficulties in using the internet and consider the patient's reading and writing skills and major languages, as well as the ways and facilities used [80]. In addition, *guide manufacturers* should design and produce special easy-to-use manuals and video tutorials for the functions of products commonly used by the older adults [51].

Research Gaps

We found that the literature included in this study still had certain research gaps. First, from the perspective of research methods, most of the included literature was based on statistical analysis and a qualitative approach to confirm the influence of various factors such as age on health inequities. Future research can use machine learning, deep learning, and other methods to explore the impact of more indicators on health inequality, dig deeper into the incidence relation between indicators, and then analyze the causal relationship of related indicators to propose more targeted countermeasures.

Second, in terms of influencing factors, scholars have studied objective factors such as age, race, rural residence, health conditions, and eHealth literacy. However, subjective factors, such as an individual's willingness to use digital health technologies and attitudes toward the internet, have not been studied. Age and education also influence eHealth literacy but the literature we have included does not study the moderating effect of age, education, and other factors on the relationship between eHealth literacy and health inequities. Therefore, future research could include subjective factors and the social support received by individuals while considering the moderating effects of multiple factors.

Finally, from the strategy perspective, the included literature showed that no more innovative technologies were applied to socially disadvantaged groups. In the new digital health era, people are demanding more high-quality and personalized health

care services. In the future, services should be provided to different socially disadvantaged groups based on their population classification. The integration of multiple digital health technologies, such as internet multiparty voice call technology and virtual reality, could be applied to disease diagnosis and treatment, virtual reality in surgery, telemedicine, and health management monitoring to assist individuals in obtaining better health care services.

Limitations

This study had several limitations. First, we only searched the literature in the three databases of Clarivate Analytics' Web of Science, PubMed, and Scopus, which may have resulted in the related literature not being retrieved. Second, owing to the large differences in the types of literature studies and outcome indicators included in the study, only 17 (41%) of the 41 included articles discussed related health consequences, such as morbidity and mortality, caused by digital health technologies. Therefore, we failed to perform a meta-analysis of quantifiable outcome measures. Third, we aimed to reduce publication bias, and although our inclusion criteria were broad, our search was limited to articles written only in English. Finally, the decision to exclude gray literature, including books and reports, might have led to the exclusion of relevant literature that could possibly have been used to widen or further support the perspectives presented in our results. Nevertheless, as gray literature includes reports and documents often drafted by political or special-interest organizations, it is more difficult to assess underlying biases, which may negatively add bias to our results.

Acknowledgments

This study was supported by the Fundamental Research Funds for the Central Universities, Huazhong University of Science and Technology (2017WKYXQY005). The authors would like to thank all anonymous reviewers for their valuable comments and inputs for this research.

Authors' Contributions

RY, the co-first author, designed the study and contributed to the data collection and manuscript writing. WZ, the co-first author, designed the study and contributed to the data collection and manuscript writing. RE, the third author, contributed to writing the manuscript and final proofreading. GC, the fourth author, contributed to the writing of the draft manuscript. TR, the fifth author, contributed to the discussion and responses of the revised manuscript. LS, the sixth and corresponding author, designed and conducted the study and finalized the draft manuscript. All the authors contributed to the preparation and approval of the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The details of PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist.

[[DOC File , 205 KB - jmir_v24i3e34144_app1.doc](#)]

References

1. Sherman LD, Grande SW. Building better clinical relationships with patients: an argument for digital health solutions with black men. *Health Serv Insights* 2019;12:1178632919834315 [[FREE Full text](#)] [doi: [10.1177/1178632919834315](https://doi.org/10.1177/1178632919834315)] [Medline: [30886522](https://pubmed.ncbi.nlm.nih.gov/30886522/)]

2. Digital strategy: leading the culture change in health and care. Department of Health. 2012. URL: <https://www.gov.uk/government/publications/digital-strategy-leading-the-culture-change-in-health-and-care> [accessed 2021-09-01]
3. State of the union: commission proposes a path to the digital decade to deliver the EU's digital transformation by 2030. European Commission. 2021. URL: https://ec.europa.eu/commission/presscorner/detail/en/ip_21_4630 [accessed 2021-09-01]
4. Draft global strategy for digital health (2020-2025). World Health Organization. 2019. URL: https://www.who.int/docs/default-source/documents/200067ec-lb-full-draft-digital-health-strategy-liu.pdf?sfvrsn=46e08cc0_2 [accessed 2021-09-01]
5. Outline of healthy China 2030 Plan. China Government Network. 2016. URL: http://www.gov.cn/zhengce/2016-10/25/content_5124174.htm [accessed 2021-09-01]
6. Rodriguez JA, Clark CR, Bates DW. Digital health equity as a necessity in the 21st Century Cures Act era. *JAMA* 2020;323(23):2381-2382. [doi: [10.1001/jama.2020.7858](https://doi.org/10.1001/jama.2020.7858)] [Medline: [32463421](https://pubmed.ncbi.nlm.nih.gov/32463421/)]
7. WHO guideline: recommendations on digital interventions for health system strengthening. World Health Organization. 2019. URL: <https://apps.who.int/iris/bitstream/handle/10665/311941/9789241550505-eng.pdf?ua=1> [accessed 2021-09-01]
8. Sharma A, Harrington RA, McClellan MB, Turakhia MP, Eapen ZJ, Steinhubl S, et al. Using digital health technology to better generate evidence and deliver evidence-based care. *J Am Coll Cardiol* 2018;71(23):2680-2690 [FREE Full text] [doi: [10.1016/j.jacc.2018.03.523](https://doi.org/10.1016/j.jacc.2018.03.523)] [Medline: [29880129](https://pubmed.ncbi.nlm.nih.gov/29880129/)]
9. Seckman C, Van de Castle B. Understanding digital health technologies using mind maps. *J Nurs Scholarsh* 2021;53(1):7-15. [doi: [10.1111/jnu.12611](https://doi.org/10.1111/jnu.12611)] [Medline: [33259152](https://pubmed.ncbi.nlm.nih.gov/33259152/)]
10. Scott BK, Miller GT, Fonda SJ, Yeaw RE, Gaudaen JC, Pavliscsak HH, et al. Advanced digital health technologies for covid-19 and future emergencies. *Telemed J E Health* 2020;26(10):1226-1233. [doi: [10.1089/tmj.2020.0140](https://doi.org/10.1089/tmj.2020.0140)] [Medline: [32456560](https://pubmed.ncbi.nlm.nih.gov/32456560/)]
11. Evidence standards framework for digital health technologies: user guide. National Institute for Health and Care Excellence (NICE). 2019. URL: <https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/user-guide.pdf> [accessed 2021-12-20]
12. Lupton D. Digitized health promotion: personal responsibility for health in the web 2.0 era. In: Davis JE, Gonzalez AM, editors. *To fix or to heal: patient care, public health, and the limits of biomedicine*. New York, NY: New York University Press; 2016:152-176.
13. Jones SS, Rudin RS, Perry T, Shekelle PG. Health information technology: an updated systematic review with a focus on meaningful use. *Ann Intern Med* 2014;160(1):48-54 [FREE Full text] [doi: [10.7326/M13-1531](https://doi.org/10.7326/M13-1531)] [Medline: [24573664](https://pubmed.ncbi.nlm.nih.gov/24573664/)]
14. Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, et al. Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. *Ann Intern Med* 2006;144(10):742-752 [FREE Full text] [doi: [10.7326/0003-4819-144-10-200605160-00125](https://doi.org/10.7326/0003-4819-144-10-200605160-00125)] [Medline: [16702590](https://pubmed.ncbi.nlm.nih.gov/16702590/)]
15. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. *Am J Prev Med* 2016;51(5):843-851 [FREE Full text] [doi: [10.1016/j.amepre.2016.06.008](https://doi.org/10.1016/j.amepre.2016.06.008)] [Medline: [27745684](https://pubmed.ncbi.nlm.nih.gov/27745684/)]
16. Berwick DM, Nolan TW, Whittington J. The triple aim: care, health, and cost. *Health Aff (Millwood)* 2008;27(3):759-769. [doi: [10.1377/hlthaff.27.3.759](https://doi.org/10.1377/hlthaff.27.3.759)] [Medline: [18474969](https://pubmed.ncbi.nlm.nih.gov/18474969/)]
17. Klein R, Huang D. Defining and measuring disparities, inequities, and inequalities in the healthy people initiative. Centers for Disease Control and Prevention. 2010. URL: https://www.cdc.gov/nchs/ppt/nchs2010/41_klein.pdf [accessed 2021-09-01]
18. McAuley A. Digital health interventions: widening access or widening inequalities? *Public Health* 2014;128(12):1118-1120. [doi: [10.1016/j.puhe.2014.10.008](https://doi.org/10.1016/j.puhe.2014.10.008)] [Medline: [25458115](https://pubmed.ncbi.nlm.nih.gov/25458115/)]
19. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. *J Am Med Inform Assoc* 2018;25(8):1080-1088 [FREE Full text] [doi: [10.1093/jamia/ocy052](https://doi.org/10.1093/jamia/ocy052)] [Medline: [29788380](https://pubmed.ncbi.nlm.nih.gov/29788380/)]
20. The World Bank report says innovation in information technology could lead to new inequities. Global Network. 2016. URL: <https://china.huanqiu.com/article/9CaKrnJTfkC> [accessed 2021-09-01]
21. Krouse HJ. COVID-19 and the widening gap in health inequity. *Otolaryngol Head Neck Surg* 2020;163(1):65-66. [doi: [10.1177/0194599820926463](https://doi.org/10.1177/0194599820926463)] [Medline: [32366172](https://pubmed.ncbi.nlm.nih.gov/32366172/)]
22. Shankardass K, Lofters A, Kirst M, Quiñonez C. Public awareness of income-related health inequalities in Ontario, Canada. *Int J Equity Health* 2012;11:26 [FREE Full text] [doi: [10.1186/1475-9276-11-26](https://doi.org/10.1186/1475-9276-11-26)] [Medline: [22613058](https://pubmed.ncbi.nlm.nih.gov/22613058/)]
23. Jennings L, Gagliardi L. Influence of mHealth interventions on gender relations in developing countries: a systematic literature review. *Int J Equity Health* 2013;12:85 [FREE Full text] [doi: [10.1186/1475-9276-12-85](https://doi.org/10.1186/1475-9276-12-85)] [Medline: [24131553](https://pubmed.ncbi.nlm.nih.gov/24131553/)]
24. Nittas V, Ameli V, Little M, Humphreys DK. Exploring the equity impact of mobile health-based human immunodeficiency virus interventions: a systematic review of reviews and evidence synthesis. *Digit Health* 2020;6:2055207620942360 [FREE Full text] [doi: [10.1177/2055207620942360](https://doi.org/10.1177/2055207620942360)] [Medline: [32742717](https://pubmed.ncbi.nlm.nih.gov/32742717/)]
25. Disadvantaged groups. European Institute for Gender Equality. 2013. URL: <https://eige.europa.eu/thesaurus/terms/1083> [accessed 2021-12-20]
26. Turnbull S, Lucas PJ, Hay AD, Cabral C. Digital health interventions for people with type 2 diabetes to develop self-care expertise, adapt to identity changes, and influence other's perception: qualitative study. *J Med Internet Res* 2020;22(12):e21328 [FREE Full text] [doi: [10.2196/21328](https://doi.org/10.2196/21328)] [Medline: [33346733](https://pubmed.ncbi.nlm.nih.gov/33346733/)]

27. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Population Health and Public Health Practice, Committee on Community-Based Solutions to Promote Health Equity in the United States. In: Baciu A, Negussie Y, Geller A, Weinstein JN, editors. *Communities in action: pathways to health equity*. Washington, D.C.: National Academies Press (US); 2017.
28. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol* 2005;8(1):19-32. [doi: [10.1080/1364557032000119616](https://doi.org/10.1080/1364557032000119616)]
29. 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;169(7):467-473 [FREE Full text] [doi: [10.7326/M18-0850](https://doi.org/10.7326/M18-0850)] [Medline: [30178033](https://pubmed.ncbi.nlm.nih.gov/30178033/)]
30. André A. The information technology revolution in health care. In: André A, editor. *Digital medicine*. Cham, Switzerland: Springer; 2019:1-7.
31. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. *Qual Health Res* 2012;22(10):1435-1443. [doi: [10.1177/1049732312452938](https://doi.org/10.1177/1049732312452938)] [Medline: [22829486](https://pubmed.ncbi.nlm.nih.gov/22829486/)]
32. Yang BM. Medical technology and inequity in health care: the case of Korea. *Health Policy Plan* 1993;8(4):385-393. [doi: [10.1093/heapol/8.4.385](https://doi.org/10.1093/heapol/8.4.385)] [Medline: [10131034](https://pubmed.ncbi.nlm.nih.gov/10131034/)]
33. Steiger TS, Duetz Schmucki M. Health information on the internet: an issue of social inequality in public health. *Soz Praventivmed* 2004;49(6):357-358. [doi: [10.1007/s00038-004-4044-4](https://doi.org/10.1007/s00038-004-4044-4)] [Medline: [15669433](https://pubmed.ncbi.nlm.nih.gov/15669433/)]
34. Viswanath K, Kreuter MW. Health disparities, communication inequalities, and eHealth. *Am J Prev Med* 2007;32(5 Suppl):S131-S133 [FREE Full text] [doi: [10.1016/j.amepre.2007.02.012](https://doi.org/10.1016/j.amepre.2007.02.012)] [Medline: [17466818](https://pubmed.ncbi.nlm.nih.gov/17466818/)]
35. Kim EH, Stolyar A, Lober WB, Herbaugh AL, Shinstrom SE, Zierler BK, et al. Challenges to using an electronic personal health record by a low-income elderly population. *J Med Internet Res* 2009;11(4):e44 [FREE Full text] [doi: [10.2196/jmir.1256](https://doi.org/10.2196/jmir.1256)] [Medline: [19861298](https://pubmed.ncbi.nlm.nih.gov/19861298/)]
36. Andreassen HK, Dyb K. Differences and inequalities in health: empirical reflections on telemedicine and politics. *Inf Commun Soc* 2010;13(7):956-975. [doi: [10.1080/1369118x.2010.499953](https://doi.org/10.1080/1369118x.2010.499953)]
37. Goldberg DS. Justice, population health, and deep brain stimulation: the interplay of inequities and novel health technologies. *AJOB Neurosci* 2012;3(1):16-20. [doi: [10.1080/21507740.2011.635626](https://doi.org/10.1080/21507740.2011.635626)]
38. Jones R. Development of a questionnaire and cross-sectional survey of patient eHealth readiness and eHealth inequalities. *Med* 2013;2(2):e9 [FREE Full text] [doi: [10.2196/med20.2559](https://doi.org/10.2196/med20.2559)] [Medline: [25075244](https://pubmed.ncbi.nlm.nih.gov/25075244/)]
39. Albright KC, Boehme AK, Mullen MT, Wu TC, Branas CC, Grotta JC, et al. The effect of telemedicine on access to acute stroke care in Texas: the story of age inequalities. *Stroke Res Treat* 2015;2015:813493 [FREE Full text] [doi: [10.1155/2015/813493](https://doi.org/10.1155/2015/813493)] [Medline: [26543664](https://pubmed.ncbi.nlm.nih.gov/26543664/)]
40. Matteucci I. Social determinants of health inequalities: moving toward a socio-constructivist model supported by information and communication technologies. *Global Bioethics* 2015;26(3-4):206-217. [doi: [10.1080/11287462.2015.1101213](https://doi.org/10.1080/11287462.2015.1101213)]
41. van Mierlo T, Hyatt D, Ching AT. Employing the Gini coefficient to measure participation inequality in treatment-focused digital health social networks. *Netw Model Anal Health Inform Bioinform* 2016;5(1):32 [FREE Full text] [doi: [10.1007/s13721-016-0140-7](https://doi.org/10.1007/s13721-016-0140-7)] [Medline: [27840788](https://pubmed.ncbi.nlm.nih.gov/27840788/)]
42. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. *J Med Internet Res* 2017;19(4):e136 [FREE Full text] [doi: [10.2196/jmir.6731](https://doi.org/10.2196/jmir.6731)] [Medline: [28450271](https://pubmed.ncbi.nlm.nih.gov/28450271/)]
43. Hosseinpoor AR, Nambiar D, Tawilah J, Schlottheuber A, Briot B, Bateman M, Suparmi, et al. Capacity building for health inequality monitoring in Indonesia: enhancing the equity orientation of country health information system. *Glob Health Action* 2018;11(sup1):1419739 [FREE Full text] [doi: [10.1080/16549716.2017.1419739](https://doi.org/10.1080/16549716.2017.1419739)] [Medline: [29569528](https://pubmed.ncbi.nlm.nih.gov/29569528/)]
44. Bol N, Helberger N, Weert JC. Differences in mobile health app use: a source of new digital inequalities? *Inf Soc* 2018;26;34(3):183-193. [doi: [10.1080/01972243.2018.1438550](https://doi.org/10.1080/01972243.2018.1438550)]
45. Weiss D, Rydland HT, Øversveen E, Jensen MR, Solhaug S, Krokstad S. Innovative technologies and social inequalities in health: a scoping review of the literature. *PLoS One* 2018;13(4):e0195447 [FREE Full text] [doi: [10.1371/journal.pone.0195447](https://doi.org/10.1371/journal.pone.0195447)] [Medline: [29614114](https://pubmed.ncbi.nlm.nih.gov/29614114/)]
46. Gann B. Transforming lives: combating digital health inequality. *IFLA J* 2019;45(3):187-198. [doi: [10.1177/0340035219845013](https://doi.org/10.1177/0340035219845013)]
47. Toscos T, Drouin M, Pater J, Flanagan M, Pfafman R, Mirro MJ. Selection biases in technology-based intervention research: patients' technology use relates to both demographic and health-related inequities. *J Am Med Inform Assoc* 2019;26(8-9):835-839 [FREE Full text] [doi: [10.1093/jamia/ocz058](https://doi.org/10.1093/jamia/ocz058)] [Medline: [31172193](https://pubmed.ncbi.nlm.nih.gov/31172193/)]
48. Qureshi S, Xiong J, Deitenbeck B. The effect of mobile health and social inequalities on human development and health outcomes: mhealth for health equity. In: *Proceedings of the 52nd Hawaii International Conference on System Sciences*. 2019 Presented at: HICCS '19; January 8-11, 2019; Grand Wailea, HI p. 3943-3952 URL: <https://scholarspace.manoa.hawaii.edu/bitstream/10125/59831/1/0391.pdf> [doi: [10.24251/HICSS.2019.478](https://doi.org/10.24251/HICSS.2019.478)]
49. Hansen AH, Bradway M, Broz J, Claudi T, Henriksen Ø, Wangberg SC, et al. Inequalities in the use of eHealth between socioeconomic groups among patients with type 1 and type 2 diabetes: cross-sectional study. *J Med Internet Res* 2019;21(5):e13615 [FREE Full text] [doi: [10.2196/13615](https://doi.org/10.2196/13615)] [Medline: [31144669](https://pubmed.ncbi.nlm.nih.gov/31144669/)]

50. Baum F, Newman L, Biedrzycki K. Vicious cycles: digital technologies and determinants of health in Australia. *Health Promot Int* 2014;29(2):349-360. [doi: [10.1093/heapro/das062](https://doi.org/10.1093/heapro/das062)] [Medline: [23144236](https://pubmed.ncbi.nlm.nih.gov/23144236/)]
51. Banerjee A. Digital health interventions and inequalities: the case for a new paradigm. *BMJ Evid Based Med* 2021;26(3):77-78. [doi: [10.1136/bmjebm-2019-111282](https://doi.org/10.1136/bmjebm-2019-111282)] [Medline: [31753847](https://pubmed.ncbi.nlm.nih.gov/31753847/)]
52. Rich E, Miah A, Lewis S. Is digital health care more equitable? The framing of health inequalities within England's digital health policy 2010-2017. *Sociol Health Illn* 2019;41 Suppl 1:31-49 [FREE Full text] [doi: [10.1111/1467-9566.12980](https://doi.org/10.1111/1467-9566.12980)] [Medline: [31599987](https://pubmed.ncbi.nlm.nih.gov/31599987/)]
53. Ahmed T, Rizvi SJ, Rasheed S, Iqbal M, Bhuiya A, Standing H, et al. Digital health and inequalities in access to health services in Bangladesh: mixed methods study. *JMIR Mhealth Uhealth* 2020;8(7):e16473 [FREE Full text] [doi: [10.2196/16473](https://doi.org/10.2196/16473)] [Medline: [32706736](https://pubmed.ncbi.nlm.nih.gov/32706736/)]
54. Glied S, Lleras-Muney A. Technological innovation and inequality in health. *Demography* 2008;45(3):741-761 [FREE Full text] [doi: [10.1353/dem.0.0017](https://doi.org/10.1353/dem.0.0017)] [Medline: [18939670](https://pubmed.ncbi.nlm.nih.gov/18939670/)]
55. Fujioka JK, Budhwani S, Thomas-Jacques T, De Vera K, Challa P, Fuller K, et al. Challenges and strategies for promoting health equity in virtual care: protocol for a scoping review of reviews. *JMIR Res Protoc* 2020;9(12):e22847 [FREE Full text] [doi: [10.2196/22847](https://doi.org/10.2196/22847)] [Medline: [33211020](https://pubmed.ncbi.nlm.nih.gov/33211020/)]
56. Khilnani A, Schulz J, Robinson L. The COVID-19 pandemic: new concerns and connections between eHealth and digital inequalities. *J Inf Commun Ethics Soc* 2020;18(3):393-403. [doi: [10.1108/jices-04-2020-0052](https://doi.org/10.1108/jices-04-2020-0052)]
57. DeGuzman PB, Siegfried Z, Leimkuhler ME. Evaluation of rural public libraries to address telemedicine inequities. *Public Health Nurs* 2020;37(5):806-811. [doi: [10.1111/phn.12777](https://doi.org/10.1111/phn.12777)] [Medline: [32715533](https://pubmed.ncbi.nlm.nih.gov/32715533/)]
58. Gann B. Combating digital health inequality in the time of coronavirus. *J Consum Health Internet* 2020;24(3):278-284. [doi: [10.1080/15398285.2020.1791670](https://doi.org/10.1080/15398285.2020.1791670)]
59. Bommakanti KK, Smith LL, Liu L, Do D, Cuevas-Mota J, Collins K, et al. Requiring smartphone ownership for mHealth interventions: who could be left out? *BMC Public Health* 2020;20(1):81 [FREE Full text] [doi: [10.1186/s12889-019-7892-9](https://doi.org/10.1186/s12889-019-7892-9)] [Medline: [31959145](https://pubmed.ncbi.nlm.nih.gov/31959145/)]
60. Weiss D, Sund ER, Freese J, Krokstad S. The diffusion of innovative diabetes technologies as a fundamental cause of social inequalities in health. The Nord-Trøndelag Health Study, Norway. *Sociol Health Illn* 2020;42(7):1548-1565. [doi: [10.1111/1467-9566.13147](https://doi.org/10.1111/1467-9566.13147)] [Medline: [32539185](https://pubmed.ncbi.nlm.nih.gov/32539185/)]
61. Karri K, Yarra P. Inequities still exist in the use of digital health technology across different sociodemographic subgroups. *Evid Based Nurs* 2022;25(1):23. [doi: [10.1136/ebnurs-2020-103355](https://doi.org/10.1136/ebnurs-2020-103355)] [Medline: [33262165](https://pubmed.ncbi.nlm.nih.gov/33262165/)]
62. Jaffe DH, Lee L, Huynh S, Haskell TP. Health inequalities in the use of telehealth in the United States in the lens of COVID-19. *Popul Health Manag* 2020;23(5):368-377. [doi: [10.1089/pop.2020.0186](https://doi.org/10.1089/pop.2020.0186)] [Medline: [32816644](https://pubmed.ncbi.nlm.nih.gov/32816644/)]
63. DeGuzman PB, Bernacchi V, Cupp CA, Dunn B, Ghamandi BJ, Hinton ID, et al. Beyond broadband: digital inclusion as a driver of inequities in access to rural cancer care. *J Cancer Surviv* 2020;14(5):643-652. [doi: [10.1007/s11764-020-00874-y](https://doi.org/10.1007/s11764-020-00874-y)] [Medline: [32390103](https://pubmed.ncbi.nlm.nih.gov/32390103/)]
64. Sun N, Esom K, Dhaliwal M, Amon JJ. Human rights and digital health technologies. *Health Hum Rights* 2020;22(2):21-32 [FREE Full text] [Medline: [33390689](https://pubmed.ncbi.nlm.nih.gov/33390689/)]
65. Ukoha EP, Feinglass J, Yee LM. Disparities in electronic patient portal use in prenatal care: retrospective cohort study. *J Med Internet Res* 2019;21(9):e14445 [FREE Full text] [doi: [10.2196/14445](https://doi.org/10.2196/14445)] [Medline: [31586367](https://pubmed.ncbi.nlm.nih.gov/31586367/)]
66. Hamideh D, Nebeker C. The digital health landscape in addiction and substance use research: will digital health exacerbate or mitigate health inequities in vulnerable populations? *Curr Addict Rep* 2020;7(3):317-332. [doi: [10.1007/s40429-020-00325-9](https://doi.org/10.1007/s40429-020-00325-9)]
67. Erhunmwunsee L, Bhandari P, Sosa E, Sur M, Ituarte PH, Lui NS. Socioeconomic, rural, and insurance-based inequities in robotic lung cancer resections. *Video-Assist Thorac Surg* 2020;5:13. [doi: [10.21037/vats.2020.02.01](https://doi.org/10.21037/vats.2020.02.01)]
68. Ramírez AS, Estrada E, Ruiz A. Mapping the health information landscape in a rural, culturally diverse region: implications for interventions to reduce information inequality. *J Prim Prev* 2017;38(4):345-362 [FREE Full text] [doi: [10.1007/s10935-017-0466-7](https://doi.org/10.1007/s10935-017-0466-7)] [Medline: [28224349](https://pubmed.ncbi.nlm.nih.gov/28224349/)]
69. Matthews KA, Gallo LC. Psychological perspectives on pathways linking socioeconomic status and physical health. *Annu Rev Psychol* 2011;62:501-530 [FREE Full text] [doi: [10.1146/annurev.psych.031809.130711](https://doi.org/10.1146/annurev.psych.031809.130711)] [Medline: [20636127](https://pubmed.ncbi.nlm.nih.gov/20636127/)]
70. Norman CD, Skinner HA. eHEALS: the eHealth literacy scale. *J Med Internet Res* 2006;8(4):e27 [FREE Full text] [doi: [10.2196/jmir.8.4.e27](https://doi.org/10.2196/jmir.8.4.e27)] [Medline: [17213046](https://pubmed.ncbi.nlm.nih.gov/17213046/)]
71. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. *JMIR Mhealth Uhealth* 2014;2(1):e2 [FREE Full text] [doi: [10.2196/mhealth.2994](https://doi.org/10.2196/mhealth.2994)] [Medline: [25098314](https://pubmed.ncbi.nlm.nih.gov/25098314/)]
72. Rogers EM. Diffusion of innovation. 4th edition. New York, NY: Free Press; 2010.
73. Campos-Castillo C. Revisiting the first-level digital divide in the United States: gender and race/ethnicity patterns, 2007-2012. *Soc Sci Comput Rev* 2015;33(4):423-439. [doi: [10.1177/0894439314547617](https://doi.org/10.1177/0894439314547617)]
74. Schradie J. The trend of class, race, and ethnicity in social media inequality: who still cannot afford to blog? *Inf Commun Soc* 2012;15(4):555-571. [doi: [10.1080/1369118x.2012.665939](https://doi.org/10.1080/1369118x.2012.665939)]

75. Link BG, Phelan J. Social conditions as fundamental causes of disease. *J Health Soc Behav* 1995;Spec No:80-94. [Medline: [7560851](#)]
76. Tichenor PJ, Donohue GA, Olien CN. Mass media flow and differential growth in knowledge. *Public Opin Q* 1970;34(2):159-170. [doi: [10.1086/267786](#)]
77. Mirowsky J, Ross CE. Education, social status, and health. New York, NY: Routledge; 2003.
78. Kocher R, Emanuel EJ, DeParle NA. The Affordable Care Act and the future of clinical medicine: the opportunities and challenges. *Ann Intern Med* 2010;153(8):536-539 [FREE Full text] [doi: [10.7326/0003-4819-153-8-201010190-00274](#)] [Medline: [20733178](#)]
79. Kontos E, Blake KD, Chou WY, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. *J Med Internet Res* 2014;16(7):e172 [FREE Full text] [doi: [10.2196/jmir.3117](#)] [Medline: [25048379](#)]
80. Apter AJ. Can patient portals reduce health disparities? A perspective from asthma. *Ann Am Thorac Soc* 2014;11(4):608-612. [doi: [10.1513/AnnalsATS.201401-032PS](#)] [Medline: [24640983](#)]

Abbreviations

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews

SES: socioeconomic status

SPIDER: Sample, Phenomenon of Interest, Design, Evaluation, Research

Edited by A Mavragani; submitted 08.10.21; peer-reviewed by S Bhattacharjee, M Tran, J Taylor, P DeGuzman, B Chaudhry; comments to author 26.11.21; revised version received 15.01.22; accepted 24.02.22; published 21.03.22.

Please cite as:

Yao R, Zhang W, Evans R, Cao G, Rui T, Shen L

Inequities in Health Care Services Caused by the Adoption of Digital Health Technologies: Scoping Review

J Med Internet Res 2022;24(3):e34144

URL: <https://www.jmir.org/2022/3/e34144>

doi: [10.2196/34144](#)

PMID: [35311682](#)

©Rui Yao, Wenli Zhang, Richard Evans, Guang Cao, Tianqi Rui, Lining Shen. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 21.03.2022. 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.

Review

Success Factors of Medical Crowdfunding Campaigns: Systematic Review

Xiaoyan Hou¹, BSc; Tailai Wu¹, PhD; Zhuo Chen^{2,3}, PhD; Liqin Zhou¹, PhD

¹School of Medicine and Health Management, Huazhong University of Science and Technology, Wuhan, China

²School of Economics, Faculty of Humanities and Social Sciences, University of Nottingham Ningbo, Ningbo, China

³Department of Health Policy and Management, College of Public Health, University of Georgia, Athens, GA, United States

Corresponding Author:

Tailai Wu, PhD

School of Medicine and Health Management

Huazhong University of Science and Technology

Hangkong Road 13

Wuhan, 430030

China

Phone: 86 13477072665

Email: lncl2012@gmail.com

Abstract

Background: Medical crowdfunding provides opportunities for individuals who lack financial resources to access the health services that they need. Despite the popularity of medical crowdfunding, the current understanding of the success of medical crowdfunding campaigns is fragmented and inadequate.

Objective: We aimed to comprehensively investigate which factors lead to the success of medical crowdfunding campaigns.

Methods: A search was conducted in PubMed, PsycINFO, Web of Science, ACM Digital Library, and ScienceDirect from 2010 to June 2020. Papers directly and indirectly related to the success of medical crowdfunding campaigns were included. Two reviewers independently extracted information on the success of medical crowdfunding campaigns.

Results: Our search yielded 441 articles, of which 13 met the inclusion criteria. Medical crowdfunding is increasingly attracting academic attention, and most studies leverage text analysis as their research methods; however, there is a lack of consensus on the definition of medical crowdfunding among researchers. Four categories of factors that affect the success of medical crowdfunding were identified: platforms, raisers, donors, and campaigns.

Conclusions: Although some limitations exist in our systematic review, our study captured and mapped literatures of the success of medical crowdfunding campaigns systematically, which can be used as the basis for future research on this topic.

(*J Med Internet Res* 2022;24(3):e30189) doi:[10.2196/30189](https://doi.org/10.2196/30189)

KEYWORDS

medical crowdfunding; success factor; systematic review; methodology; theories; preconceptions; crowdfunding; fundraising; financial resources; health care

Introduction

Crowdfunding has a massive impact on how people access health care services [1]. Due to the low requirements and easy set-up of crowdfunding websites, the usage of crowdfunding platforms for health-related fundraising is growing increasingly popular. Crowdfunding platforms, such as GoFundMe, Kickstarter, and FundRazr, are highly utilized for raising funds for a variety of causes, especially medical needs [2]. Some crowdfunding websites report that the category *medical campaigns* ranks as the top-grossing [3]. The usage of

crowdfunding websites for medical expenses is expected to increase by 25% annually [3]. Web-based medical crowdfunding emerged after 2008, during which the economic crunch led to an equally enthusiastic increase in the use and accessibility of social media platforms [4]. Medical crowdfunding is donation-based and is used to raise money for those who have medical costs. Medical crowdfunding may represent how people respond to the gaps in national health payment systems [5]; many health needs that are not met by national health insurance coverage are reflected in medical crowdfunding platforms [5-7]. The lower the national insurance coverage, the greater the

number of medical crowdfunding projects [8]. The benefits of medical crowdfunding include expanding funder participation in the health market, improving the access to financial support, drawing funding to neglected health issues, and improving social engagement [9]. Medical crowdfunding also has been shown to reduce the rate of personal bankruptcy [10]; therefore, it is becoming an important way to deal with medical financial issues.

Generally, the success of a medical crowdfunding campaign could be defined as the degree to which the medical crowdfunding campaign achieves or exceeds the goals set by fundraisers. However, despite the convenience and popularity of medical crowdfunding websites, the success rates of medical crowdfunding campaigns on different platforms vary dramatically. For example, in China, campaigns have been reported to achieve just 18% of their goals [11], whereas, in the United States, medical crowdfunding campaigns have achieved over 40% of their goals on average [4]. Overall, only 10% of medical crowdfunding campaigns have been reported to reach their fundraising target [4], and some campaigns reach their fundraising targets within a short time, while others struggle to raise their target amount. Success factors of medical crowdfunding campaigns are factors that lead the campaigns to achieve or exceed the target amount [12]. Thus, knowing these factors could help improve the success rate of medical crowdfunding campaign.

Medical crowdfunding can not only make up for the deficiencies of health insurance systems but can also address problems such as limited financing channels and low private capital utilization rates [1,13,14]. However, the negative consequences of medical crowdfunding have also become apparent and have been increasing [3,15-17]. Factors such as low-entry barriers to launch or donate to campaigns, too much separation between raisers and donors, and anonymity could increase the risk of fraud in medical crowdfunding campaigns [18,19]. In addition, the current understanding of medical crowdfunding success factors is limited. The importance and potential issues of medical crowdfunding have stimulated the interest of academic researchers [20]. Although the success factors of medical crowdfunding campaigns have been previously considered, each study has examined the factors from different perspectives and using samples from different countries. For example, Durand

et al [12] analyzed success factors only from text features. Thus, current literature on the success factors of medical crowdfunding campaigns is segmented, lacking the ability to give a holistic understanding of the factors. A systematic review [21] was needed to systematically search, critically appraise, and synthesize studies, to explore the success factors of medical crowdfunding campaigns. We aimed to comprehensively and systematically investigate the factors leading to the successes of medical crowdfunding campaigns.

Methods

Overview

We chose to conduct a systematic review, rather than choosing another method, for the following reasons: First, a systematic review could provide a comprehensive understanding concerning the success factors of medical crowdfunding campaigns. The comprehensive understanding could be a solid basis for further studies on this topic. Second, a systematic review could limit bias in identifying and rejecting bias by using explicit methods. Third, conclusions of the systematic review may be more reliable and accurate because it summarizes previous literature systematically. Fourth, the findings of systematic review could be used to reduce the delay between scientific discoveries and implementation. The generalizability of findings could be established by comparing studied success factors of medical crowdfunding campaigns [22].

Literature Search

We used PubMed, PsycINFO, Web of Science, ACM Digital Library, and ScienceDirect databases. These databases were chosen because they cover most disciplines that study medical crowdfunding, namely, medicine, information, psychology, global health, computer science, and business economics. Keywords and synonyms used in this search revolved around 2 concepts—crowdfunding, and health (Textbox 1). Database searches were conducted on June 8, 2020 by 2 authors (LZ and XH) with Library and Information Science expertise. The results were exported to Zotero (version 5.0.0.0; Corporation for Digital Scholarship) and organized into folders by database. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) was used to document the review process was used.

Textbox 1. Search strategy. The terms were searched in keyword, Boolean/phrase search modes, and all fields, within PubMed, PsycINFO, Web of Science, ACM Digital Library, and ScienceDirect respectively

- 11 crowdfunding or crowd funding
- 22 health or disease or illness or medical or hospital or treatment
- 33 1 and 2

Inclusion and Exclusion Criteria

Papers that focused on the success of medical crowdfunding or other factors related to the success of crowdfunding (eg, narrative strategies and ethical factors of medical crowdfunding) published between 2010 and 2020 (because medical crowdfunding emerged after 2008) within conference proceedings or journals were included, and papers without

research design details or results or not be written in English were excluded.

Quality Assessment

Papers were rated for quality using the McGill Mixed Methods Appraisal Tool version 2018 [23]. Its internal reliability, usability, and content validity have been verified in several studies [24,25]. Quality criteria are applied based on the study

design (qualitative research, quantitative research, mixed method research) and methodology (such as randomized controlled, nonrandomized and descriptive studies). We used criteria [23] for quantitative descriptive studies: (1) Is the sampling strategy relevant to address the quantitative research question? (2) Is the sample representative of the population under study? (3) Are measurements appropriate? (4) Is there an acceptable response rate? We also used a quality threshold (75%, ie, meeting at least 3 criteria) defined by a previous study [26]. Papers that did not meet the threshold were excluded.

Data Analysis

Metadata were extracted and listed in a spreadsheet (Excel, Microsoft Inc): author region, publication date, data sources, research questions, data collection settings, methods, results, discussions, conclusions, and bibliographies. To interpret our findings, all authors discussed how to present our findings systematically. Since the conception of medical crowdfunding is dynamic, we discuss the conception, then summarize bibliographic and study information. The success factors are

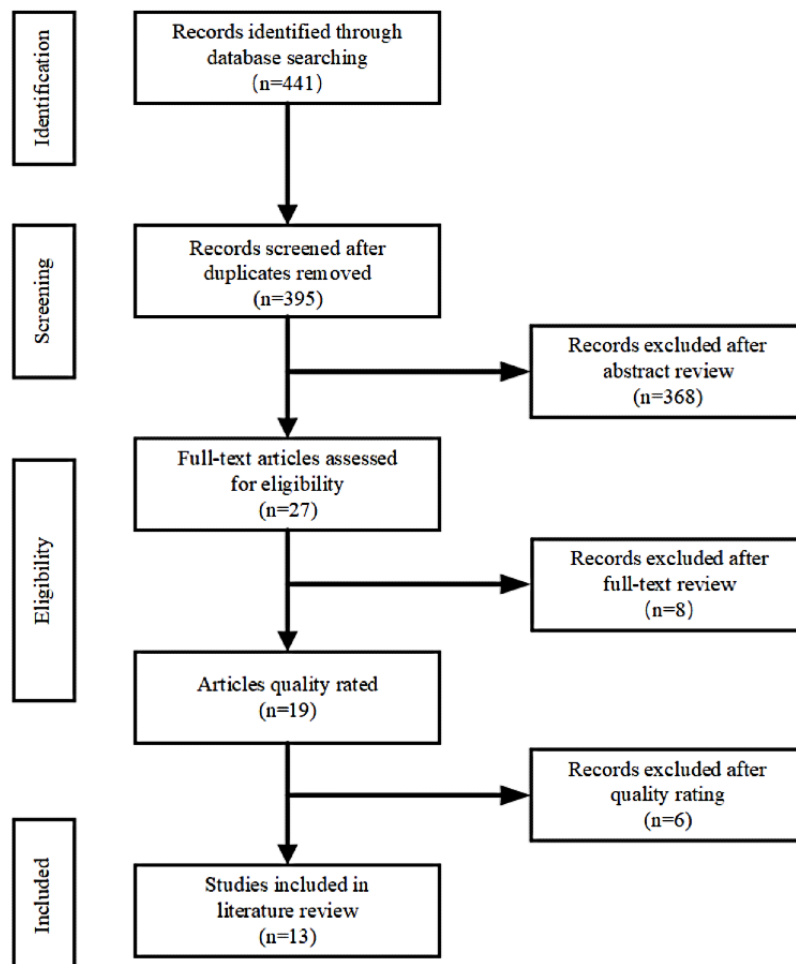
based to the contextual structure of medical crowdfunding. Since there are 3 main actors in medical crowdfunding, and they interact with each other based on specific campaigns [7], we categorized the success factors by platforms, raisers, donors, and campaigns. Because of the heterogeneity of outcome measures in studies included in our review, it was impractical to conduct a meta-analysis. Therefore, findings were qualitative.

Results

Search Results

A total of 441 papers were identified from the 5 databases. After removing 46 duplicates, titles and abstracts were screened by 2 authors separately. When a difference in opinion arose, a third author was involved to mediate the discussion to reach agreement. Using inclusion and exclusion criteria, 27 papers were chosen for full-text review. After full-text review, 19 papers were assessed for quality, 13 papers were included (Figure 1).

Figure 1. Summary of literature search and study selection process.



Medical Crowdfunding Preconceptions

Despite the acknowledgment of the benefit of medical crowdfunding in strengthening health care, there is still a lack of consensus on what constitutes crowdfunding campaigns in health care. However, to study the success factors of medical

crowdfunding, we need to have a basic understanding of medical crowdfunding. Based on previous studies [6-8], we define the scope of medical crowdfunding as crowdfunding for medical expenses in hospitals, new drugs in scientific research institutions, and new treatments (eg, hospitalization expenses, scientific research funds).

Description of the Studies

Of the 13 studies (Table 1), 9 studies [4,8,12,27-32] relied entirely on crowdfunding websites such as GoFundMe. There were 4 main analytical methods used: text analysis, regression analysis, semistructured interviews, and exploratory spatial analysis. Text analysis was used most, with 4 studies using only text analysis [4,29-31], and 2 studies using text analysis partially [12,33]. Of the 13 studies, 6 studies were directly related to the

success of medical crowdfunding, and the remaining 7 studies were indirectly related. The number of publications appears to be steadily increasing over time, which indicates that this field is getting more and more attention from scholars and practitioners because of the particularity of medical crowdfunding [6,20]. Most studies were from high-income countries, with the largest source of articles being the United States [4,12,27,32,33], followed by Canada [30,34].

Table 1. Study information.

Reference	Publication date (year, month)	Country	Study aim	Data sources	Method	Success factors
Durand et al [12]	2018,6	United States	To identify the factors influencing the success of crowdfunding campaigns	YouCaring	Text analysis, multiple linear analysis, logistic regression analysis	Campaign description length; goal amount; third-person description perspective; cognitive state
Kim et al [27]	2018,4	United States	To investigate how beneficiaries present their situations and how contributors view the information presented	GoFundMe, YouCaring, Fundly	Semistructured interviews	Authenticity
van Duynhoven et al [34]	2019,6	Canada	To explore the role of socioeconomic status in medical crowdfunding campaigns	Cancer-related activities published by Canadians; The 2016 Census Profile for aggregate dissemination area and area boundaries; forward sortation area boundaries	Exploratory spatial analysis	Socioeconomic status; demographic
Xu and Wang [28]	2019,8	China	Make clear the narrative strategy of medical crowdfunding article	Easy Fundraising	Thematic narrative analysis	Narrative strategies
Kim et al [33]	2016,5	United States	To assess the credibility of web-based medical crowdfunding campaigns	(1) Comment on Reddit related to the medical crowdfunding campaign, (2) 20 participants	Text analysis and semistructured interviews	Credibility; individual prestige
Aleksina et al [35]	2019,7	France	To investigate the determinants of successful crowdfunding campaigns in medical research	Consano, Experiment	Ordinary least square regression	Number of tweeters; goal amount; platform availability; total campaign number; total fundraising amount; total donor number
Holmes et al [29]	2019,4	United Kingdom	To determine whether crowdfunding of pharmacy-related products through popular web-based platforms	Kickstarter, Indiegogo	Text analysis	Media attention; platform audit; demographic information of donors
Snyder et al [30]	2017,6	Canada	To explore how Canadians can demonstrate to others that they should fund their health needs	FundRazr, Generosity, GoFundMe, YouCaring	Text analysis	Personal connections; depth of need; giving back; ethics
Koole et al [31]	2018,12	Netherlands	Identify key factors for the success of crowdfunding for grown-up congenital heart patients	A web-based donation platform	Text analysis	Professional organization support; stakeholder support; easy-to-understand message
Barcelos and Budge [32]	2019,1	United States	Investigated how transgender communities utilize crowdfunding expenses related to gender affirming medical care	GoFundMe	Hierarchical multiple regression analyses	Social media (Facebook) sharing; demographic information of raiser (age, location, race, identity)
Berliner et al [4]	2017,8	United States	Explore the usage, impacts, or consequences of the increasing reliance on crowdfunding for health	GoFundMe	Text analysis	Medical literacy; media literacy
Bassani et al [8]	2019,8	Italy	Examine the worldwide population of health care crowdfunding platforms and explore the relationship between health care crowdfunding success and national health systems	76 crowdfunding platforms that host health care campaigns	Negative binomial regressions	Platform type; social return

Reference	Publication date (year, month)	Country	Study aim	Data sources	Method	Success factors
Kenworthy [36]	2019,11	Norway	Map and document how medical crowdfunding is shaped by, and shapes, health disparities	An ethnography of US medical crowd-funding; a study of global health crowdfunding; a project of US medical crowd-funding campaigns	Exploratory conceptual and empirical analysis	Platform design (Search engine, lists, webpage, etc); Partnership with traditional media; Deservingness; Narratives

Success Factors

Platforms

For platforms, the factors influencing the success of medical crowdfunding campaigns can be divided into 2 aspects: technical and social (Table 2).

The *technical aspect* reflects the functionality of medical crowdfunding platform in terms of platform audit, platform availability, platform types and platform design. *Platform audit* is the review of campaigns by platforms and could impact donor’s decision on donation. Review of campaigns by

crowdfunding platforms is closely related to the eventual success of the crowdfunding campaigns [2]. *Platform availability* is the degree to which a medical crowdfunding platform is available to their users. The availability of the platform positively affects donor incentive, and thus, project funding results [36]. In addition, the type of platform can also have impact on fundraising. Platforms with more extensive publicity are more popular than specialized, smaller platforms [8]. Platform design affects donors’ experiences—supportive behavior and word of mouth helps fundraisers attract the attention of potential donors [36].

Table 2. Factors influencing the success of medical crowdfunding campaigns from the platforms.

Dimensions and factors	Definitions	Functions	Reference
Technical			
Platform audit	Review of campaigns on crowdfunding platforms	Reviews of campaigns by crowdfunding platforms affect donor decisions.	Holmes et al [29]
Platform availability	The degree to which the platform is available to users	The availability of the platform affects donor incentives, and thus, campaign funding results.	Aleksina et al [35]
Platform types	Whether the platform is specialized or general	Platforms with more extensive publicity are more popular than specialized smaller platforms.	Bassani et al [8] Aleksina et al [35]
Platform design	Design elements of the platform including search engine, lists, webpage, etc	Platform design determines donors’ experiences, and consequently, their donation behavior.	Kenworthy [36]
Social			
Total campaign number	Total number of campaigns initiated in the platform	With more categories and projects, more potential donors visit the platform.	Aleksina et al [35]
Total fundraising amount	Total money raised in the platform	Higher total amounts raised on the platform represent higher recognition and acceptance of the platform.	Aleksina et al [35]
Total donor number	Total number of donors appears in the platform	With more donors in the platform, there is a better possibility of getting funding.	Aleksina et al [35] Barcelos and Budge [32]
Partnership with traditional media	Platform collaboration with traditional media to disseminate the information of its campaigns	Traditional media can help medical crowdfunding campaigns get more donors.	Kenworthy [36]

The *social aspect* reflects interactions inside and outside the platform in terms of total campaign number, total fundraising amount, total donor number and total and partnership with traditional media. The larger the number of categories and campaigns contained, the larger the number of potential donors attracted by the platform, and the greater the probability of receiving donations [37]. If the total amount of fundraising on the platform is higher, fundraisers who publish campaigns on the platform have more confidence that they will raise the

amount they want [35]. Moreover, to some extent, the total number of donors can represent the amount that could be raised [38]. The number of registered institutions in the platform also reflects the platform’s position within the industry and the resources that are available to campaigns in the platform [3]. In addition, partnerships with traditional media allow campaigns to have opportunities to access mass media and their audiences, increasing the number of potential donors. With more interactions, there is more social capital, with higher fundraising

possibilities [39]. The larger the potential donor base, the higher the possibility of reaching the target amount [9].

Raisers

The influencing factors from raisers (ie, the beneficiaries) can be analyzed in terms of 3 aspects: demographic, individual characteristics, and social (Table 3).

The demographic characteristics of the raisers include age, nationality, and geographic locations. Younger individuals were more likely to succeed in health crowdfunding [32]. Location was also an important factor—raisers in remote areas are less successful in raising money, while raisers in affluent areas, in there are more social resources, are more likely to succeed in raising money [37]. Moreover, raisers with higher levels of education and higher income were shown to be more likely to attract the attention of potential donors and thus get donations [34].

Successful crowdfunding campaigners had better media literacy and medical literacy [4]. Raisers with high media literacy are more likely to spread their message across social media platforms and attract potential donors. Raisers with higher medical literacy are more likely to provide accurate disease-related and health care information [4]. In addition, once raisers are perceived to be deserving, their campaigns attract many donors and succeed in raising funds [36].

Social factors included personal connections, stakeholder support, professional organization support, and individual prestige. Raisers’ personal connections establish social networks and can share links, which allows more potential donors to see the information of the campaigns, thus obtaining more funds. Gaining the support of stakeholders and professional organizations makes it easier to successfully raise funds for medical crowdfunding campaigns [31], and if raisers have high individual prestige, they are more likely to gain trust and support from donors in social networks [30].

Table 3. Factors influencing the success of medical crowdfunding campaigns from the raisers.

Dimensions and factors	Definitions	Functions	Sources
Demographic			
Age	Age	Younger people are more likely to succeed in health crowdfunding.	Aleksina et al [35]; Barcelos and Budge [32]
Education level	Level of education	Raisers with higher levels of education are more likely to attract the attention of donors (higher donation possibility).	van Duynhoven et al [34]
Income	Income	Raisers with higher income are more likely to attract the attention of donors (higher donation possibility).	van Duynhoven et al [34]
Geographical location	Where raisers reside	Raisers are more likely to get help from people in or near their districts, especially in wealthier places.	Barcelos and Budge [32]
Individual characteristics			
Media literacy	Raisers’ ability to make use of different media	The raisers of successful crowdfunding campaigns have good media literacy.	Berliner and Kenworthy [4]; Holmes et al [29]
Medical literacy	Raisers’ ability of leverage different medical knowledge	A certain level of medical literacy (of raisers) facilitates the proper description of the disease and relevant understanding of the health care system.	Berliner et al [4]
Deservingness	The degree to which raisers are thought to be deserving of receiving donations	Once raisers are perceived to be deserving, their campaigns attract many donors, and they succeed in raising funds.	Kenworthy [36]
Social			
Personal connections	Raisers’ personal connections with others including their families, friends, and colleagues	The scale of raisers’ personal connections has a positive effect on the success rate of fundraising.	Snyder et al [30]
Stakeholder and professional organization support	Raisers get support from different stakeholders and professional organizations	The support from stakeholders and professional organizations makes fundraising easier.	Koole et al [31]
Individual prestige	Raisers’ personal respect and admiration from others inside and outside the platform	The prestige of raisers can serve as the signal of the credibility and success of their campaigns.	Kim et al [33]; Snyder et al [30]

Donors

The determinants of intention to donate in donors can be mainly divided into demographic and individual characteristics (Table 4).

Donor gender [40,41], age [29], education level [29], income [29], and geographical location [35] had significant effects on donation behaviors. Younger donors were more willingness to donate [32], and people with higher education and income level were more likely to donate [42]. Donors were more willing to contribute to crowdfunding campaigns in or near their own regions [35]; therefore, geographic inequity exists, which is compounded by social, technological, and cultural issues [4].

Individual characteristics included cognitive state and social returns. *Cognitive state* is state invoked in the donor in reading a campaign's description [12]. If donors feel threatened by reading the negative campaign descriptions, they hesitate to donate. When donors are aware the importance and urgency of the raiser's medical crowdfunding campaigns, they have higher willingness to donate [43]. *Social return* is a major intrinsic motivation for individuals or groups to donate to medical crowdfunding campaigns [4]. People with prosocial values and who participate more in charitable activities donate more willingly [44]. Donors with intrinsic motivation have high willingness to donate.

Table 4. Factors influencing the success of medical crowdfunding campaigns from donors.

Dimensions and factors	Definitions	Functions	Sources
Demographic			
Age	Age	Younger donors have higher willingness to donate.	Holmes et al [29]
Education level	Level of education	Donors with higher levels of education have higher willingness to donate.	Holmes et al [29]
Income	Income	Donors with higher income have higher willingness to donate.	Holmes et al [29]
Geographical location	Where raisers reside	Donors are more willing to contribute to crowdfunding campaigns in or near their own regions.	Aleksina et al [35]; Berliner and Kenworthy [4]
Individual characteristics			
Cognitive state	The cognitive state of the donor when they read the campaign	Donors may feel threatened by negative campaign descriptions and social pressure to donate. The positive cognitive state invoked by the campaign description would promote donors' donation behavior.	Durand et al [12]
Social returns	The intrinsic motivation of donors to give back to society	The more returns to the society from the donation, the more possibility donors would donate.	Bassani et al [8]

Campaigns

Success factors for campaigns can be divided into 2 aspects: the format and the content (Table 5).

Format-related factors include goal amount, campaign description length, third-person description, and social media sharing. When a campaign is close to its fundraising goal, it encourages donation behavior and increases the likelihood of project success [40,45]. Longer campaign descriptions and higher goal amount were significantly associated with amount raised [12]. In addition, third-person perspective in the description can convey a patient's positive qualities in a way that would be not acceptable in the first-person perspective due to the testimonial effect [12]. It is also critical for crowdfunding campaigns to leverage social media [35]. The amount raised was strongly correlated with updates and shares in social media [12]. One additional tweet or retweet with more personal comments could enhance the probability of success of crowdfunding campaigns [35].

Content-related factors included the narrative strategy, authenticity, credibility, being easy to understand, giving back,

and depth of need. *Illness narratives* represent the personal story and illness experience that patients are sharing verbally or in writing [28]. The narration style not only affected the efficiency of the dissemination of health information but also heavily affected on the potential donors' cognitions, attitudes, and behaviors [28]. Authenticity, which can be conveyed by pictures of raisers that depict their medical conditions, increases the possibility of donations. Funding is more accessible when the presented narratives of campaigns were credible [46]. To demonstrate the credibility of campaigns, many methods, including collective endorsements, presenting details of external financial support, displaying off-site verification details (of ailment, incident, and treatment), providing verification of fundraiser and beneficiary identities, and using a popular and trusted platform, can be used [33].

Easy-to-understand information in the campaign description gives potential donors a clear picture of the campaign [31]. In addition, portraying the beneficiary as someone who selflessly gives back to society not only helps establish the positive image of raisers, but also, inspires the donors themselves [30]. Patients in urgent need of funds because of disease or for treatment are easy to obtain donations from potential donors [30].

Table 5. Factors influencing the success of medical crowdfunding campaigns from campaigns.

Dimensions and factors	Definitions	Functions	Sources
Format			
Goal amount	The objective amount of money the campaign plan to raise.	Goal amount has a positive impact on campaign success.	Durand et al [12]; Aleksina et al [35]
Campaign description length	The length of the medical crowdfunding campaign description.	Campaign description length has a positive impact on campaign success.	Durand et al [12]
Third-person description perspective	The narrative perspective of the campaign is the third person.	The third-person perspective makes the story more objective and realistic, which makes it more convincing.	Durand et al [12]
Social media sharing	The number of shares and likes of campaigns in social media which connect to the platform.	The more shares and likes potential donors see, the more likely they are to donate.	Barcelos and Budge [32]
Content			
Narrative strategy	The way that raisers describe their illness and ask for donations.	Narrative strategies such as more positive emotions, more information, and appropriate arousal level have impacts on crowdfunding success.	Durand et al [12]; Koole et al [31]; Kenworthy [36]
Authenticity	The content of medical crowdfunding campaigns is authentic.	Potential donors pay attention on their impression of raisers' authentic medical conditions and make their decisions based on it.	Kim et al [27]
Credibility	Credibility of medical crowdfunding campaigns.	Credibility of campaigns which could be formed based collective endorsement have impacts on the success of medical crowdfunding campaigns.	van Duynhoven et al [34]; Kim et al [33]; Koole et al [31]
Easy-to understand message	The degree to which the message is easily understood.	Easy-to-understand information helps potential donors get a sense of the raisers' intention, which in turn helps donors make decisions.	Koole et al [31]
Giving back	Portraying the raisers as someone who selflessly gives back to society.	The past efforts of the raisers on behalf of others were used as a rationale for the potential donors to contribute to the crowdfunding campaign.	Snyder et al [30]
Deep of need	Campaign content reflects the urgent need of funds to solve health problems.	The urgency of need for help would determine the success of raising money.	Snyder et al [30]

Discussion

Principal Findings

By investigating 13 studies, the key factors of successful medical crowdfunding campaigns were extracted. On this basis, we conducted a more in-depth review and provide a comprehensive understanding of the factors that influence successful campaigns. We find the success factors of medical crowdfunding campaigns could be divided into 4 categories: platforms, raisers, donors, and campaigns. The success factors involve the main actors in the whole medical crowdfunding campaigns, but campaign factors were more frequently studied than other factors. Platform factors, however, played an important role in medical crowdfunding [47], because platforms are important in linking donors and raisers [48]. There were relatively few studies related to donors; thus, future studies could pay more attention to this area. Despite the widespread use of medical crowdfunding, to the best of our knowledge, this is the first systematic review that examines the factors that lead to successful medical crowdfunding campaigns.

Implications

The 4 categories of successful factors that are actors in medical crowdfunding interact with each other [7,49,50]. First, raisers register in the platforms to initiate campaigns to raise money for medical expenses with illness narratives, which include fundraising goals, fundraising time, and fundraising events. Second, the medical crowdfunding platforms conduct a preliminary review to release the campaigns. Third, managers of the platforms give the funds raised from the public to the campaigners after the relevant processing charges are collected. A crowdfunding campaign will be success if all the above steps are were successfully implemented.

Theory for understanding the success factors of medical crowdfunding can be developed from our findings. None of articles included in this review take strong theoretical perspective to help analyze, explain, and predict the success factors of medical crowdfunding and convey few theoretical implications. Therefore, it is necessary to apply more theories to better understand this topic. Second, more methods or mixed methods could be employed to investigate this topic. Although text analysis was used most often, more insights could be explored using other methods that possess different advantages.

We found that data were collected from single sources, which may result inherent biases; collecting data from multiple sources could allow validation to proposed research models and used methods in different studies.

Limitations and Future Directions

Our study has some limitations. First, no new factors and relationships about medical crowdfunding could be explored through systematic review. This was the most significant inherent limitation of our study since we relied on previously published literature. Second, additional literature could have been included in our systematic review. Literature written in other languages from other databases may also have discussed the success factors of medical crowdfunding. Third, bias may

still exist in the process of selection and evaluation of literature. Although we attempted to minimize bias from differences between reviewers, such as solving conflicts in judgment by discussion among reviewers, bias cannot be excluded. Fourth, heterogeneity exists in the literature included in the review; therefore, the quality of our review results may be impacted. Other categories of success factors, such as health systems or national economic status, could be also be considered.

Conclusion

By examining platform, raiser, donor, and campaign-related success factors, we provide information that can be used as the basis for future research and future medical crowdfunding campaigns.

Acknowledgments

This work was partially supported by grants from the National Natural Science Foundation of China (grants 71801100, 72004072, and 72004071), the Humanities and Social Science Foundation of Ministry of Education of China (grants 20YJC870016), and the Fundamental Research Funds for the Central Universities (grant 2019kfyXJJS172).

Conflicts of Interest

None declared.

References

1. Young MJ, Scheinberg E. The rise of crowdfunding for medical care: promises and perils. *JAMA* 2017 Apr 25;317(16):1623-1624. [doi: [10.1001/jama.2017.3078](https://doi.org/10.1001/jama.2017.3078)] [Medline: [28334321](https://pubmed.ncbi.nlm.nih.gov/28334321/)]
2. Zenone M, Snyder J. Fraud in medical crowdfunding: a typology of publicized cases and policy recommendations. *Policy Internet* 2018 Oct 28;11(2):215-234. [doi: [10.1002/poi3.188](https://doi.org/10.1002/poi3.188)]
3. Snyder J. Crowdfunding for medical care: ethical issues in an emerging health care funding practice. *Hastings Cent Rep* 2016 Nov 22;46(6):36-42. [doi: [10.1002/hast.645](https://doi.org/10.1002/hast.645)] [Medline: [27875643](https://pubmed.ncbi.nlm.nih.gov/27875643/)]
4. 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](https://doi.org/10.1016/j.socscimed.2017.02.008)] [Medline: [28274601](https://pubmed.ncbi.nlm.nih.gov/28274601/)]
5. Snyder J, Zenone M, Crooks V, Schuurman N. What medical crowdfunding campaigns can tell us about local health system gaps and deficiencies: exploratory analysis of British Columbia, Canada. *J Med Internet Res* 2020 May 22;22(5):e16982 [FREE Full text] [doi: [10.2196/16982](https://doi.org/10.2196/16982)] [Medline: [32348269](https://pubmed.ncbi.nlm.nih.gov/32348269/)]
6. Lublóy Á. Medical crowdfunding in a healthcare system with universal coverage: an exploratory study. *BMC Public Health* 2020 Nov 09;20(1):1672 [FREE Full text] [doi: [10.1186/s12889-020-09693-3](https://doi.org/10.1186/s12889-020-09693-3)] [Medline: [33167927](https://pubmed.ncbi.nlm.nih.gov/33167927/)]
7. 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](https://doi.org/10.2196/18813)] [Medline: [32618573](https://pubmed.ncbi.nlm.nih.gov/32618573/)]
8. Bassani G, Marinelli N, Vismara S. Crowdfunding in healthcare. *J Technol Transf* 2018 Apr 20;44(4):1290-1310. [doi: [10.1007/s10961-018-9663-7](https://doi.org/10.1007/s10961-018-9663-7)]
9. Renwick MJ, Mossialos E. Crowdfunding our health: economic risks and benefits. *Soc Sci Med* 2017 Oct;191:48-56 [FREE Full text] [doi: [10.1016/j.socscimed.2017.08.035](https://doi.org/10.1016/j.socscimed.2017.08.035)] [Medline: [28889030](https://pubmed.ncbi.nlm.nih.gov/28889030/)]
10. Burtch G, Chan J. Investigating the Relationship Between Medical Crowdfunding and Personal Bankruptcy in the United States: Evidence of a Digital Divide. *MIS Q* 2019 Jan 1;43(1):237-262. [doi: [10.25300/misq/2019/14569](https://doi.org/10.25300/misq/2019/14569)] [Medline: [Investigating the relationship between medical crowdfunding and personal bankruptcy in the United States: evidence of a digital divide](https://pubmed.ncbi.nlm.nih.gov/32618573/)]
11. Jin P. Medical crowdfunding in China: empirics and ethics. *J Med Ethics* 2019 Aug;45(8):538-544. [doi: [10.1136/medethics-2018-105054](https://doi.org/10.1136/medethics-2018-105054)] [Medline: [31217228](https://pubmed.ncbi.nlm.nih.gov/31217228/)]
12. Durand WM, Peters JL, Eltorai AEM, Kalagara S, Osband AJ, Daniels AH. Medical crowdfunding for organ transplantation. *Clin Transplant* 2018 Jun;32(6):e13267. [doi: [10.1111/ctr.13267](https://doi.org/10.1111/ctr.13267)] [Medline: [29683220](https://pubmed.ncbi.nlm.nih.gov/29683220/)]
13. Sisler J. Crowdfunding for medical expenses. *Can Med Assoc J* 2012 Feb 07;184(2):E123-E124 [FREE Full text] [doi: [10.1503/cmaj.109-4084](https://doi.org/10.1503/cmaj.109-4084)] [Medline: [22231688](https://pubmed.ncbi.nlm.nih.gov/22231688/)]
14. Pol SJ, Snyder J, Anthony SJ. "Tremendous financial burden": crowdfunding for organ transplantation costs in Canada. *PLoS ONE* 2019 Dec 20;14(12):e0226686. [doi: [10.1371/journal.pone.0226686](https://doi.org/10.1371/journal.pone.0226686)]
15. Dressler G, Kelly SA. Ethical implications of medical crowdfunding: the case of Charlie Gard. *J Med Ethics* 2018 Jul;44(7):453-457. [doi: [10.1136/medethics-2017-104717](https://doi.org/10.1136/medethics-2017-104717)] [Medline: [29728452](https://pubmed.ncbi.nlm.nih.gov/29728452/)]

16. Simon R. Crowdfunding sites Like GoFundMe and YouCaring raise money and concerns. *Wall Street Journal*. 2016 Mar 1. URL: <https://www.wsj.com/articles/crowdfunding-sites-like-gofundme-and-youcaring-raise-moneyand-concerns-1456775949> [accessed 2020-12-22]
17. Robinson L, Cotten SR, Ono H, Quan-Haase A, Mesch G, Chen W, et al. Digital inequalities and why they matter. *Inf Commun Soc* 2015 Mar 16;18(5):569-582. [doi: [10.1080/1369118X.2015.1012532](https://doi.org/10.1080/1369118X.2015.1012532)]
18. Siering M, Koch J, Deokar A. Detecting fraudulent behavior on crowdfunding platforms: the role of linguistic and content-based cues in static and dynamic contexts. *J Manag Inf Syst* 2016 Oct 05;33(2):421-455 [FREE Full text] [doi: [10.1080/07421222.2016.1205930](https://doi.org/10.1080/07421222.2016.1205930)]
19. Snyder J, Cohen IG. Medical crowdfunding for unproven medical treatments: should Gofundme become a gatekeeper. *Hastings Cent Rep* 2019 Nov;49(6):32-38. [doi: [10.1002/hast.1066](https://doi.org/10.1002/hast.1066)] [Medline: [31813183](https://pubmed.ncbi.nlm.nih.gov/31813183/)]
20. Kenworthy N, Dong Z, Montgomery A, Fuller E, Berliner L. A cross-sectional study of social inequities in medical crowdfunding campaigns in the United States. *PLoS One* 2020 Mar 5;15(3):e0229760 [FREE Full text] [doi: [10.1371/journal.pone.0229760](https://doi.org/10.1371/journal.pone.0229760)] [Medline: [32134984](https://pubmed.ncbi.nlm.nih.gov/32134984/)]
21. Tawfik GM, Dila KAS, Mohamed MYF, Tam DNH, Kien ND, Ahmed AM, et al. A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Trop Med Health* 2019;47:46 [FREE Full text] [doi: [10.1186/s41182-019-0165-6](https://doi.org/10.1186/s41182-019-0165-6)] [Medline: [31388330](https://pubmed.ncbi.nlm.nih.gov/31388330/)]
22. Pocock SJ. Systematic reviews. *BMJ* 1995 Sep 16;311(7007):759-759. [doi: [10.1136/bmj.311.7007.759](https://doi.org/10.1136/bmj.311.7007.759)]
23. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The mixed methods appraisal tool (MMAT) version 2018 for information professionals and researchers. *EFI* 2018 Dec 18;34(4):285-291. [doi: [10.3233/EFI-180221](https://doi.org/10.3233/EFI-180221)]
24. Pace R, Pluye P, Bartlett G, Macaulay AC, Salsberg J, Jagosh J, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *Int J Nurs Stud* 2012 Jan;49(1):47-53. [doi: [10.1016/j.ijnurstu.2011.07.002](https://doi.org/10.1016/j.ijnurstu.2011.07.002)] [Medline: [21835406](https://pubmed.ncbi.nlm.nih.gov/21835406/)]
25. Clifford BK, Mizrahi D, Sandler CX, Barry BK, Simar D, Wakefield CE, et al. Barriers and facilitators of exercise experienced by cancer survivors: a mixed methods systematic review. *Support Care Cancer* 2018 Mar;26(3):685-700. [doi: [10.1007/s00520-017-3964-5](https://doi.org/10.1007/s00520-017-3964-5)] [Medline: [29185105](https://pubmed.ncbi.nlm.nih.gov/29185105/)]
26. 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](https://doi.org/10.1016/j.ijmedinf.2018.07.007)] [Medline: [30153920](https://pubmed.ncbi.nlm.nih.gov/30153920/)]
27. Kim JG, Hong H, Karahalios K. Understanding identity presentation in medical crowdfunding. 2018 Apr Presented at: CHI Conference on Human Factors in Computing Systems; April 21-26, 2018; Montreal, Canada p. 1-12. [doi: [10.1145/3173574.3173708](https://doi.org/10.1145/3173574.3173708)]
28. 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](https://doi.org/10.1080/10410236.2019.1654173)] [Medline: [31455114](https://pubmed.ncbi.nlm.nih.gov/31455114/)]
29. Holmes TM, Aungst TD, Smith CC, Metcalf MD. Crowdfunding pharmacy- and medication-related products: how successful is it? *J Am Pharm Assoc (2003)* 2019;59(2S):S57-S62. [doi: [10.1016/j.japh.2019.01.002](https://doi.org/10.1016/j.japh.2019.01.002)] [Medline: [30772207](https://pubmed.ncbi.nlm.nih.gov/30772207/)]
30. Snyder J, Crooks VA, Mathers A, Chow-White P. Appealing to the crowd: ethical justifications in Canadian medical crowdfunding campaigns. *J Med Ethics* 2017 Jun;43(6):364-367 [FREE Full text] [doi: [10.1136/medethics-2016-103933](https://doi.org/10.1136/medethics-2016-103933)] [Medline: [28137998](https://pubmed.ncbi.nlm.nih.gov/28137998/)]
31. Koole MAC, Kauw D, Winter MM, Schuurin MJ. A successful crowdfunding project for eHealth research on grown-up congenital heart disease patients. *Int J Cardiol* 2018 Dec 15;273:96-99 [FREE Full text] [doi: [10.1016/j.ijcard.2018.10.015](https://doi.org/10.1016/j.ijcard.2018.10.015)] [Medline: [30297187](https://pubmed.ncbi.nlm.nih.gov/30297187/)]
32. Barcelos CA, Budge SL. Inequalities in crowdfunding for transgender health care. *Transgend Health* 2019 May;4(1):81-88 [FREE Full text] [doi: [10.1089/trgh.2018.0044](https://doi.org/10.1089/trgh.2018.0044)] [Medline: [31032423](https://pubmed.ncbi.nlm.nih.gov/31032423/)]
33. Kim J, Kong H, Karahalios K, Fu W, Hong H. The power of collective endorsements: credibility factors in medical crowdfunding campaigns. 2016 Presented at: Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, California. [doi: [10.1145/2858036.2858289](https://doi.org/10.1145/2858036.2858289)]
35. van Duynhoven A, Lee A, Michel R, Snyder J, Crooks V, Chow-White P, et al. Spatially exploring the intersection of socioeconomic status and Canadian cancer-related medical crowdfunding campaigns. *BMJ Open* 2019 Jun 20;9(6):e026365 [FREE Full text] [doi: [10.1136/bmjopen-2018-026365](https://doi.org/10.1136/bmjopen-2018-026365)] [Medline: [31227531](https://pubmed.ncbi.nlm.nih.gov/31227531/)]
35. Aleksina A, Akulenkina 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](https://doi.org/10.1016/j.drudis.2019.05.012)] [Medline: [31132416](https://pubmed.ncbi.nlm.nih.gov/31132416/)]
36. Kenworthy NJ. Crowdfunding and global health disparities: an exploratory conceptual and empirical analysis. *Global Health* 2019 Nov 28;15(Suppl 1):71 [FREE Full text] [doi: [10.1186/s12992-019-0519-1](https://doi.org/10.1186/s12992-019-0519-1)] [Medline: [31775896](https://pubmed.ncbi.nlm.nih.gov/31775896/)]
37. Burtch G, Ghose A, Wattal S. Cultural differences and geography as determinants of online prosocial lending. *MIS Q* 2014 3;38(3):773-794. [doi: [10.25300/misq/2014/38.3.07](https://doi.org/10.25300/misq/2014/38.3.07)]
38. Belleflamme P, Lambert T, Schwienbacher A. Crowdfunding: tapping the right crowd. *J Bus Venturing* 2014 Sep;29(5):585-609. [doi: [10.1016/j.jbusvent.2013.07.003](https://doi.org/10.1016/j.jbusvent.2013.07.003)]
39. Gleasure R, Feller J. Does heart or head rule donor behaviors in charitable crowdfunding markets? *Int J Electro Commerce* 2016 Jun 20;20(4):499-524. [doi: [10.1080/10864415.2016.1171975](https://doi.org/10.1080/10864415.2016.1171975)]

40. Ortiz RA, Witte S, Gouw A, Sanfilippo A, Tsai R, Fumagalli D, et al. Engaging a community for rare genetic disease: best practices and education from individual crowdfunding campaigns. *Interact J Med Res* 2018 Apr 27;7(1):e10707 [FREE Full text] [doi: [10.2196/10707](https://doi.org/10.2196/10707)] [Medline: [30578239](https://pubmed.ncbi.nlm.nih.gov/30578239/)]
41. Su H, Chou T, Osborne PG. When financial information meets religion: charitable-giving behavior in Taiwan. *Soc Behav Pers* 2011 Sep 01;39(8):1009-1019. [doi: [10.2224/sbp.2011.39.8.1009](https://doi.org/10.2224/sbp.2011.39.8.1009)]
42. Dunn P. Professional corporate donation programs in Canada: an exploratory study. *Nonprofit Volunt Sect Q* 2016 Jun 30;33(2):334-345. [doi: [10.1177/0899764004263422](https://doi.org/10.1177/0899764004263422)]
43. Bekkers R, Wiepking P. A literature review of empirical studies of philanthropy: eight mechanisms that drive charitable giving. *Nonprofit Volunt Sect Q* 2010 Sep 10;40(5):924-973. [doi: [10.1177/0899764010380927](https://doi.org/10.1177/0899764010380927)]
44. Van Lange PAM, Bekkers R, Schuyt TNM, Vugt MV. From games to giving: social value orientation predicts donations to noble causes. *Basic Appl Soc Psychol* 2010 Jun 15;29(4):375-384. [doi: [10.1080/01973530701665223](https://doi.org/10.1080/01973530701665223)]
45. 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](https://doi.org/10.1016/j.jbusvent.2013.06.005)]
46. Herzenstein M, Sonenshein S, Dholakia UM. Tell me a good story and i may lend you my money: the role of narratives in peer-to-peer lending decisions. SSRN 2011:S138-S149. [doi: [10.2139/ssrn.1840668](https://doi.org/10.2139/ssrn.1840668)]
47. 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 May 04;23(5):e25554 [FREE Full text] [doi: [10.2196/25554](https://doi.org/10.2196/25554)] [Medline: [33944787](https://pubmed.ncbi.nlm.nih.gov/33944787/)]
48. Huang Z, Ouyang J, Huang X, Yang Y, Lin L. Explaining donation behavior in medical crowdfunding in social media. *SAGE Open* 2021 Apr 30;11(2):215824402110145. [doi: [10.1177/21582440211014520](https://doi.org/10.1177/21582440211014520)]
49. Swan M. Crowdsourced health research studies: an important emerging complement to clinical trials in the public health research ecosystem. *J Med Internet Res* 2012 Mar 07;14(2):e46 [FREE Full text] [doi: [10.2196/jmir.1988](https://doi.org/10.2196/jmir.1988)] [Medline: [22397809](https://pubmed.ncbi.nlm.nih.gov/22397809/)]
50. Lee D, Park J. The relationship between a charity crowdfunding project's contents and donors' participation: an empirical study with deep learning methodologies. *Comput Human Behav* 2020 May;106:106261. [doi: [10.1016/j.chb.2020.106261](https://doi.org/10.1016/j.chb.2020.106261)]

Abbreviations

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

Edited by T Leung; submitted 04.05.21; peer-reviewed by M Lukk, J Zhou; comments to author 28.06.21; revised version received 11.07.21; accepted 26.02.22; published 22.03.22.

Please cite as:

Hou X, Wu T, Chen Z, Zhou L

Success Factors of Medical Crowdfunding Campaigns: Systematic Review

J Med Internet Res 2022;24(3):e30189

URL: <https://www.jmir.org/2022/3/e30189>

doi: [10.2196/30189](https://doi.org/10.2196/30189)

PMID: [35315779](https://pubmed.ncbi.nlm.nih.gov/35315779/)

©Xiaoyan Hou, Tailai Wu, Zhuo Chen, Liqin Zhou. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 22.03.2022. 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.

Review

Process and Outcome Evaluations of Smartphone Apps for Bipolar Disorder: Scoping Review

Iona Tatham^{1*}; Ellisiv Clarke^{1*}, MBBS; Kelly Ann Grieve^{1,2}, BA; Pulkit Kaushal^{1,2}, MBBS, MD; Jan Smeddinck³, PhD; Evelyn Barron Millar¹, PhD; Aditya Narain Sharma^{1,2}, MBBS, MD, PhD

¹Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, United Kingdom

²National Specialist Adolescent Mood disorders Service, Cumbria Northumberland Tyne and Wear NHS Foundation Trust, Walkergate Park, Newcastle upon Tyne, United Kingdom

³Open Lab, Human Computer Interaction, Urban Sciences Building, Newcastle University, Newcastle upon Tyne, United Kingdom

*these authors contributed equally

Corresponding Author:

Aditya Narain Sharma, MBBS, MD, PhD

Translational and Clinical Research Institute

Faculty of Medical Sciences

Newcastle University

Academic Psychiatry, Wolfson Research Centre

Campus for Ageing and Vitality

Newcastle upon Tyne, NE4 5PL

United Kingdom

Phone: 44 1912875262

Email: aditya.sharma@ncl.ac.uk

Abstract

Background: Mental health apps (MHAs) provide opportunities for accessible, immediate, and innovative approaches to better understand and support the treatment of mental health disorders, especially those with a high burden, such as bipolar disorder (BD). Many MHAs have been developed, but few have had their effectiveness evaluated.

Objective: This systematic scoping review explores current process and outcome measures of MHAs for BD with the aim to provide a comprehensive overview of current research. This will identify the best practice for evaluating MHAs for BD and inform future studies.

Methods: A systematic literature search of the health science databases PsycINFO, MEDLINE, Embase, EBSCO, Scopus, and Web of Science was undertaken up to January 2021 (with no start date) to narratively assess how studies had evaluated MHAs for BD.

Results: Of 4051 original search results, 12 articles were included. These 12 studies included 435 participants, and of these, 343 had BD type I or II. Moreover, 11 of the 12 studies provided the ages (mean 37 years) of the participants. One study did not report age data. The male to female ratio of the 343 participants was 137:206. The most widely employed validated outcome measure was the Young Mania Rating Scale, being used 8 times. The Hamilton Depression Rating Scale-17/Hamilton Depression Rating Scale was used thrice; the Altman Self-Rating Mania Scale, Quick Inventory of Depressive Symptomatology, and Functional Assessment Staging Test were used twice; and the Coping Inventory for Stressful Situations, EuroQoL 5-Dimension Health Questionnaire, Generalized Anxiety Disorder Scale-7, Inventory of Depressive Symptomatology, Mindfulness Attention Awareness Scale, Major Depression Index, Morisky-Green 8-item, Perceived Stress Scale, and World Health Organization Quality of Life-BREF were used once. Self-report measures were captured in 9 different studies, 6 of which used MONARCA. Mood and energy levels were the most commonly used self-report measures, being used 4 times each. Furthermore, 11 of the 12 studies discussed the various confounding factors and barriers to the use of MHAs for BD.

Conclusions: Reported low adherence rates, usability challenges, and privacy concerns act as barriers to the use of MHAs for BD. Moreover, as MHA evaluation is itself developing, guidance for clinicians in how to aid patient choices in mobile health needs to develop. These obstacles could be ameliorated by incorporating co-production and co-design using participatory patient approaches during the development and evaluation stages of MHAs for BD. Further, including qualitative aspects in trials that

examine patient experience of both mental ill health and the MHA itself could result in a more patient-friendly fit-for-purpose MHA for BD.

(*J Med Internet Res* 2022;24(3):e29114) doi:[10.2196/29114](https://doi.org/10.2196/29114)

KEYWORDS

child and adolescent mental health; scoping review; bipolar disorder; mental health

Introduction

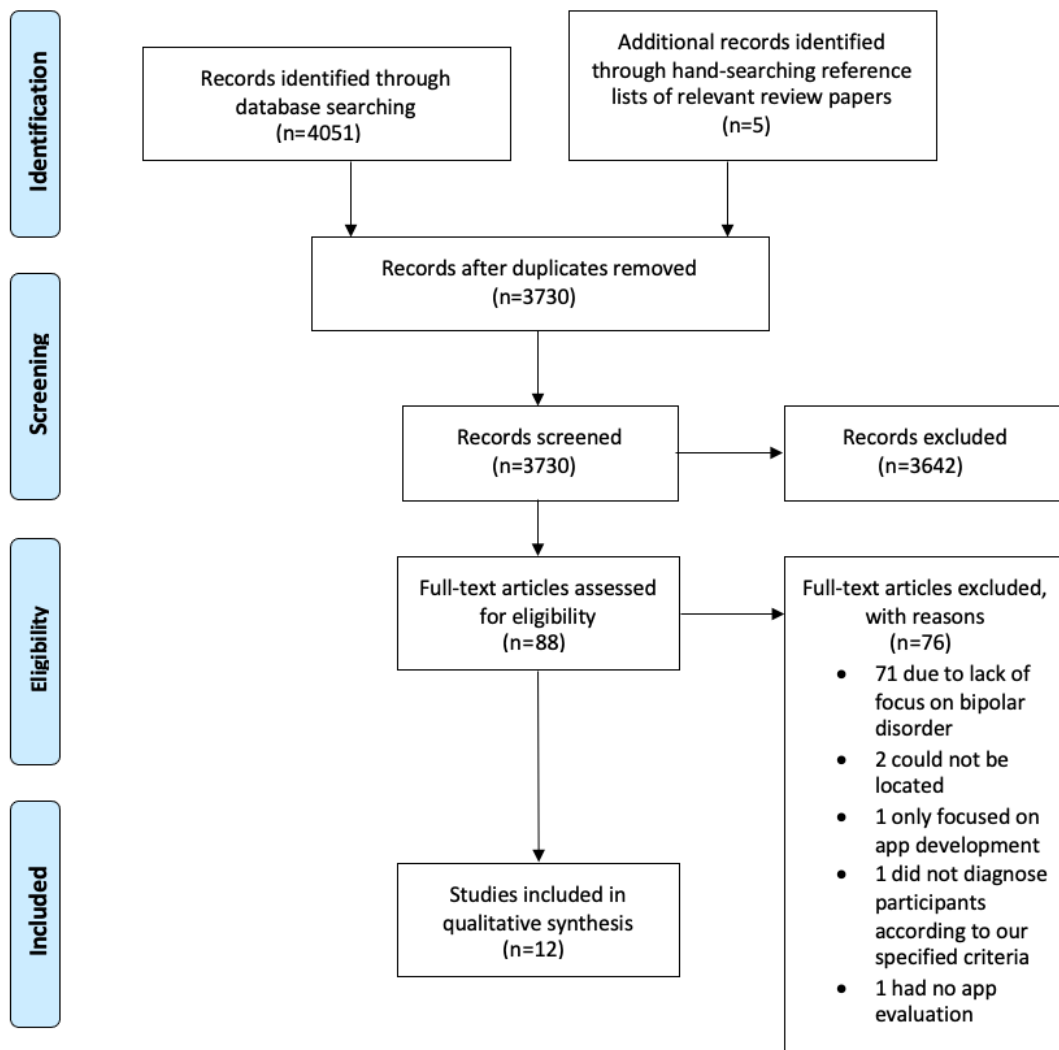
Background

There are many critical factors that can influence the course and outcome of mental health disorders. Two key factors are (1) early and accurate identification of the first onset and subsequent relapses of the disorder, leading to the institution of appropriate management, and (2) access to appropriate treatment locations. For bipolar disorder (BD), the average delay between the onset of symptoms and the first institution of treatment can be as long as 10 to 15 years [1-3]. Between 35% and 50% of patients with mental health disorders receive no treatment because appropriate treatment locations are rare [4]. BD is no exception to this rule. A United Kingdom-based 2015 study found that the median diagnostic delay in the South

London and Maudsley National Health Service (NHS) Trust was 62 days, with the median treatment delay being a further 31 days [5]. Research regarding pathways to redress these delays is urgently required, and with the potential to reliably scaffold processes and scale to both large numbers and remote locations, digital technology holds considerable potential to address these challenges.

In 2020, an estimated 6 billion smartphones were in use across the globe [6]. In the United Kingdom, there has been a 79% increase in the number of 5- to 15-year-old children owning mobile phones since 2015 [7]. Although smartphone ownership tends to be more common in high-income countries, as economies develop, the price of smartphones will decrease, and this correlation will reduce [6]. One form of digital technology that can capitalize on this increased smartphone usage globally is mental health apps (MHAs).

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



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Prior Work

Currently, MHAs can be seen to improve engagement and accessibility for individuals in rural areas where health care provision is increasingly difficult to access [8,9], and they are well accepted by service users [10]. So much so that financial incentives have been implemented for behavioral health information technologies in US policy [11]. Such advances in digital mental health care have now been adopted by clinicians in the treatment of common mental disorders in the United Kingdom [9]. Cost-effectiveness is crucial to health care, especially in a government-funded system as comprehensive as the NHS. Evidence suggests that the use of tele-psychiatry interventions reduced pressure on mental health services in low-

and middle-income countries in comparison to a control group [12], but noted the importance of rigorous app evaluation. This is echoed by Tal et al [13], who described both the potential opportunities and risks posed by digital mental health, and how they must be balanced in order to achieve meaningful change.

The socioeconomic cost of BD in the United Kingdom is well recognized [14]. Previous literature suggests that the use of MHAs for BD can increase patient engagement and provide real-time symptom monitoring to allow for improved recognition of symptoms of relapse [15]. Subsequently, this reduces barriers to treatment, such as lack of resources and time. However, the efficacy of MHAs for BD is unclear [16], and a paucity of evidence in how to assess and evaluate MHAs for BD makes these statements difficult to qualify. To date, there is a lack of

regulatory guidelines regarding MHAs, including those for BD, as health technologies are a relatively new resource within the NHS. This could potentially lead to unsafe use and practice [17]. Little is known about how MHAs (including those for BD) are developed and scrutinized, and studies predict that consumers, policymakers, health services, and funders will demand a robust evaluation process before funding, prescribing, or using these services [18].

As the development of an MHA for BD requires iterative processes with stakeholders outside of the academic and clinical research environment, process evaluation is important (in addition to more traditional outcome measure methods) to ensure the app remains user-friendly and functional without compromising clinical outcomes. This scoping review aims to address this research gap through mapping existing literature on process and outcome evaluation methods of MHAs for BD to increase the understanding around currently available evaluation tools and the latest practice.

Aim

The aim of this scoping review is to systematically explore current process and outcome measures to identify the best practice for evaluating MHAs for BD. The focus is on apps for BD designed for individuals across the lifespan. Conducting a scoping review will allow health care systems to be more structurally informed on how to accurately evaluate the effectiveness of such technologies when implementing them into routine care [19]. The specific objectives of this scoping review, based on the detailed framework of Levac et al [20], were to map the available evidence and report on (1) process evaluation methods (ie, participant usability and functionality), (2) outcome measure methods (ie, data on how widely a measure is used and concordance of the target population with completing a measure), (3) outcome measures used to measure mental health improvement (eg, well-being measures), and (4) methods for best practice in the evaluation of MHAs for BD.

Methods

Overview

The review was informed by the Arksey and O'Malley 5-step framework [19], which was further developed by the Levac, Colquhoun, and O'Brien model [20]. This includes identification of a research question, study selection and criteria, data extraction, and content analysis. Employing this methodological framework will support in examining the broader field of the evaluation of MHAs for BD to identify the best practice.

Search Strategy

A scoping search was initially performed in the following databases: PsycINFO, MEDLINE, Embase, EBSCO, Scopus, and Web of Science. Relevant search terms were identified from key papers, and the search strategy was developed iteratively in MEDLINE and then translated across the other databases, up to January 2021. Due to time constraints, grey literature sources were not investigated, and the search was limited to articles published in English, as no resources were available to undertake translation work. Broad search terms were used to reduce the likelihood of article omission. The complete search

strategy for MEDLINE is available in [Multimedia Appendix 1](#). The reference lists of included studies were hand searched for additional reports of relevance.

Selection Criteria

For studies to be included in this review, they needed to meet the following inclusion and exclusion criteria. Studies were included if they (1) were related to BD, (2) targeted individuals across the lifespan, (3) included qualitative and/or quantitative evaluation methods and measures, (4) were published in the English language, (5) had no start date limit, (6) included any type of study design, (7) included participants with symptoms of BD or diagnosed with BD according to International Classification of Diseases-10, Diagnostic and Statistical Manual of Mental Disorders-IV, or Diagnostic and Statistical Manual of Mental Disorders-5, (8) included evaluation of the functionality of the MHA and/or evaluation of the participant outcomes, and (9) included any function (eg, screening, mood monitoring, or medication adherence). Studies were excluded if they (1) were based on a web-based intervention with no MHA counterpart, (2) included MHAs that were only psychotherapeutic intervention specific with no evaluation, (3) were based on MHA development only, and (4) included a participant population without symptoms of BD or not diagnosed with BD. Where systematic review papers were identified, these were not included. However, their reference lists were hand searched to identify primary articles relevant for inclusion.

Given that the aim of the study was to recognize the scope of research already conducted, both qualitative and quantitative research designs were included. As few studies focused on children and adolescents as their participant population, no age limits were applied.

Selection Process

The search was completed by 2 researchers (IT and PK), who independently screened articles by the title and abstract against the inclusion criteria. Articles that fulfilled the inclusion criteria were then subjected to full-text screening by IT and PK. Conflicts were discussed with a third researcher (EBM) to reach consensus. Eight conflicts arose altogether, including 6 when screening the title, 1 when screening the title and abstract, and 1 when screening the full paper.

Data Collection Process

Characterization of the data and the results were exported into a customized data extraction form that was piloted in a subset of included studies. Data extracted included study name, authors, year, country, study design, MHA for BD, whether the MHA for BD was independent or adjunctive, sample size, mean age of the participants, gender of the participants, inclusion/exclusion criteria for the participants, results, tools used, measures used, time points, and whether it addressed any of the 4 objectives.

Data Synthesis and Quality Assessment

EC and IT analyzed the data using narrative synthesis and placed this in the context of the current literature to formulate conclusions. The studies were assessed using the Mixed Methods Appraisal Tool (MMAT) [21].

Results

Overview

The original database search yielded 4051 articles. Hand searching of relevant review articles was conducted, which yielded a further 5 articles. After duplicates were removed, 3730 articles remained. Screening of the title and abstract resulted in 3642 articles being excluded. The remaining 88 articles were then subjected to full-text screening, and 76 articles were excluded (71 due to a lack of focus on BD, 2 could not be located, 1 only focused on app development, 1 did not diagnose participants according to our specified criteria, and 1 had no app evaluation).

Study Characteristics

Overall, 12 studies were identified as part of this review, which evaluated 7 MHAs for BD. [Multimedia Appendix 2](#) describes the characteristics and assessment compliance of each study [22-33], and [Multimedia Appendix 3](#) describes the results of the respective compliance with the standards set out in the MMAT [22-33]. Across all 12 studies, data from 435 participants were analyzed (343 with BD). Five out of the 12 studies stated the type of BD for 167 participants (112 had bipolar I disorder, 52 had bipolar II disorder, and 3 had bipolar disorder not otherwise specified). Eleven of the 12 studies provided the mean age (37 years) of the participants with BD. All 12 studies provided information on the gender (M:F of 137:206) of the participants with BD.

Assessment of the quality of all 12 studies (quantitative and mixed methods) was performed (IT and EBM) using the MMAT (version 2018) [21]. The results of their respective compliance with the standards set out in the MMAT can be found in [Multimedia Appendix 3](#). Scores ranged from 20% to 100%. However, low-quality studies were not excluded in order to summarize the small pool of available literature.

Process Evaluation

Six studies examined the self-perceived participant usability and functionality of the MHAs for BD. This examination ranged from detailed feedback questionnaires given to the participants [22,23] to participant feedback suggesting that a reminder prompt from the MHA for BD increased completion rates. Only 1 study mentioned functionality problems in MHAs for BD. Bardram et al [23] commented that MONARCA only worked

63 out of the 69 days of the study period and the information quality score was lower due to unresponsive error messages. The authors also noted that the Android market locked the app during the study period, negatively affecting the pattern of usage during that time.

Two studies recognized that technical problems were likely to arise and so implemented a system to solve these problems. Hidalgo-Mazzei et al [34] supplied participants with technical support via telephone, so they could contact the researcher for further assistance. A similar system was put in place by Schärer et al [24]. Subjects were able to report errors and receive immediate assistance by phone or personal communication. However, it was found that the MHA for BD required a certain amount of knowledge as a prerequisite, which restricted its use in comparison to a text message equivalent [24].

Outcome Evaluation

A variety of validated outcome measures were used to evaluate the selected MHAs for BD. The most widely employed measure was the Young Mania Rating Scale (n=8) [35]. The Hamilton Depression Rating Scale [36] was applied 3 times, and the Altman Self-Rating Mania Scale [37], Quick Inventory of Depressive Symptomatology [38], and Functional Assessment Staging Test [39] were used on 2 occasions. The Coping Inventory for Stressful Situations [40], EuroQoL 5-Dimension Health Questionnaire [41], Generalized Anxiety Disorder Scale-7 [41], Inventory of Depressive Symptomatology [42], Mindfulness Attention Awareness Scale [43], Major Depression Index [44], Morisky-Green 8-item [45], Perceived Stress Scale [46], and World Health Organization Quality of Life-BREF [47] were all utilized just once. Other measures were assessed in 9 different studies, 6 of which used MONARCA. These measures included mood, sleep length, medication taken, activity level, irritability, mixed mood, cognitive problems, alcohol consumption, stress level, menstruation, individualized early warning sign, energy level, anxiety, elation, sadness, anger, speed of thoughts, and impulsivity. Mood and energy level were the most commonly utilized measures, being used 4 times each.

Outcome Measure Methods

Eleven of the 12 studies presented a debate on the confounding factors affecting the efficacy of MHAs for BD. These confounding factors and the number of times they were mentioned in the 11 studies are shown in [Table 1](#).

Table 1. Identified potential confounders for mental health app efficacy.

Potential confounder	Number of studies in which mentioned
Participants were mainly stable or euthymic	4
Participant insight when experiencing a manic phase varies	3
Sample size was too small	3
Length of study was too short	3
Low retention or adherence rate	2
Method of objective data collection was not robust enough	2
Patients were found to be capable of experiencing both manic and depressive symptoms concurrently	1
Questionnaires given were too simplistic	1
Opportunity of free-text input not given	1
Error in the app	1
Change in mobile phone communication habits	1
Low prevalence of manic symptoms	1
Potential sampling bias	1
Order of questions in the questionnaire did not vary and so was open to mindless input	1
Scales not delivered often enough	1
Participants switched the mental health app (MHA) off during the study	1
Participants may have chosen not to complete the surveys due to their mood	1
Subjective scales	1
The MHA gave daily confrontation with depressive symptoms	1
The MHA was not sensitive enough to manic or depressive symptoms	1
Participants were already involved in a medication adherence intervention	1

Evaluation of MHAs for BD

Only 5 of the 12 studies commented on the future of the evaluation of MHAs for BD. Streicher et al [48] suggested that instead of measuring relapse or recurrence of affective episodes, a more sensitive measure would be assessment of mood instability or subsyndromal symptoms. They also commented that future research should include patients with bipolar disorder not otherwise specified, as this patient group may represent a large proportion of patients with BD. Hidalgo-Mazzei et al [34] acknowledged the low retention and adherence rates of MHAs for BD, and stated that researchers should focus on developing new approaches to motivate and engage patients with the intervention in the long term. The authors suggested adopting a user-centered design approach or incorporating gamification elements into a formal psychoeducation process.

Osmani et al [25] commented on the personalization of MHAs for BD, with a focus on physical activity; however, the authors found it difficult to generalize their results to the wider population. This was due to substantial variations between patients for both overall physical activity levels and physical activity levels within daily intervals. Therefore, the authors considered that an adaptive approach to user modeling would be better suited to detect early warning signs of the onset of episodes of BD and facilitate timely intervention. This involves personalizing goals and achievements around each patient's

individual needs. This has been evidenced in previous conference proceedings [48].

Schwartz et al [26] found that the generalizability of the results was limited due to a lack of a comparison group with a differing psychiatric diagnosis, which may exhibit overlapping symptoms. They recommended that future research should use additional comparison groups to better differentiate between symptoms.

Faurholt-Jepsen et al [27] proposed that emphasis should be placed on the differentiation between day-to-day difficulties and depressive symptoms. A positive reinforcing feedback mechanism may help minimize the negative processing bias and so, in theory, relieve the sustained depressive symptoms. They addressed the notion that it can be difficult for an intervention to have an effect on both depressive and manic symptoms, given the complexity of BD [27]. These suggestions are in keeping with the existing literature [49].

Discussion

Principal Findings and Comparison With Prior Work

The aim of this scoping review was to better understand how MHAs for BD are being evaluated, particularly in terms of the process of use and outcome measures. Due to the scarcity of studies evaluating MHAs for BD specifically, inferences for discussion have been assumed from studies evaluating general

MHAs. This relies on the assumption that the functions are similar.

The need for effective and diligent evaluation of MHAs for BD is well established in the literature; no credentialing is currently required for their development and release. Karcher et al [50] warned that the “questionable content” and sparse evidence base of the myriad of current MHAs available warrant careful consideration. Effective robust evaluation systems would be required in aiding patients and practitioners in making individualized appropriate decisions regarding their role and treatment options in patient care. The NHS in the United Kingdom made considerable progress in this area when they launched their Digital Technology Assessment Criteria (DTAC) in February 2021. This provides a “simpler and faster assessment process to help give staff, patients, and the public confidence that the digital health tools they use meet NHS standards” [51]. The DTAC bring together legislation and good practice into a core document [52] that all digital health technologies have to meet in order to be recommended by clinical policy teams within NHS England and NHS Improvement. Though this goes so far to provide the public with centrally regulated technologies, clinicians may lack the knowledge and skills required for the effective recommendation of an MHA [25-27,43-54]. Mindfulness and meditation MHAs are the most commonly recommended by general practitioners in Australia [54]. However, the clinical presentation of BD and its specialist management may deter general practitioners from researching or recommending MHAs for its monitoring or management. Therefore, training health care professionals’ skills in identifying and selecting high-quality MHAs for BD would be beneficial for patients. If, as the literature suggests, the use of MHAs decreases service use [10], the financial benefit may outweigh the cost of the additional training required.

One barrier in the development of MHA evaluation systems is the adherence rate. O’Connell [55] reported that 74% of users stop engaging with an MHA after only 10 uses. Low adherence reduces the confidence with which researchers can generalize their results [56]. Aforementioned personalization and gamification of MHAs for BD have been recommended; however, engagement can tail off once the initial novelty effect of the feature has subsided. It has been suggested that a change in the communication approach may help to solve this problem. Kenny et al [57] surmised that it is possible that in studies where participants (specifically young people) are aware of the importance of their input to achieve the research objective, engagement levels may be higher. This brings into focus the importance of participatory co-design and co-production of MHAs for BD. Eight of our 12 studies mentioned adherence rates (adherence rates were not applicable in 2 studies, and another 2 studies failed to mention the rates), with the average rate being 84%. In fact, Tsana et al [28] experienced 91% adherence over the first 3 months of the study period and 81.9% in the following 9 months. The variability between the studies by O’Connell [55] and Tsanas et al [28] illustrates that there is still work to be done to achieve successful and reliable compliance with such apps.

Interestingly, one reason for the low utilization of MHAs for BD may be decreased motivation, which is often a key feature

of depressive episodes [57]. Previous literature suggests that “communities of practice” around an MHA can improve long-term concordance, with social interaction and communal use (whether in person or digitally), encouraging users to continue to access it [49,50]. Integrating a “days since last updated” screening tool would help identify early relapse in the usage of MHAs for BD, and aid in assessing clinical usability [18].

Torous et al [58] interviewed adolescent patients to identify which factors would be useful to bear in mind for MHA development. The results included safety, engagement, functionality, social interaction, awareness, gender, and participative engagement by young people. One study strongly suggested the abandonment of randomized controlled trials as a method of evaluating and improving apps, and instead called for iterative participatory research or single case designs [59]. As such, MHA developers would work in collaboration with patients throughout the design and development process in order to gain regular ecologically valid feedback so that relevant appealing prototypes are established. This could take the form of consumer-used tools or accreditation portals [60]. Then, when pilot and nonpilot studies are performed, both qualitative and quantitative data could be obtained in order to receive valuable feedback in how to further improve the app. The role of randomized controlled trials can then be established in validating the MHAs at later stages. Torous et al [58] theorized further reasons for low engagement, including poor usability, lack of a user-centric design, privacy concerns, and lack of trust. Another study found that MHA efficiency, effectiveness, memorability, and learnability and cognitive load were major usability barriers to continuing MHA use [61]. This evidence lends further strength to the argument that streamlining of the usability of MHAs should be at the core of future iterative development stages of MHAs in order to increase adherence rates and improve the ecological validity and reliability of evaluations.

Torous et al [58] also recognized accessibility as a factor to consider when developing MHAs. The Office for National Statistics stated that in 2018, 10% of the UK adult population were internet “nonusers” [62], meaning they had never used the internet or had not used it in the last 3 months. This brings the idea of the digital divide into the spotlight and shows the complexities it brings with it along with merely accessibility. The digital divide cannot be solved by just providing patients with devices, as they will also need digital literacy skills to use the device to its full potential. Even though Torous et al [58] interviewed adolescents, their study does illustrate the need to consider the skill level of the intended audience. Ennis et al [63] found that lack of technological skills was the reason for nonengagement with computers and mobile devices. Furthermore, Ennis et al found that only a quarter of their 121 participants reported familiarity and easy access to smartphones. Therefore, throughout the MHA evaluation process, patients’ skill sets, in addition to access, should be taken into consideration.

As well as employing participative engagement in MHA development, improving user awareness can come from creative measures, such as describing or advertising the MHA

appropriately. Over a quarter of MHAs for depression failed to mention depression in the title or description [64].

Moving to the user-identified priority of safety [29,58], third parties obtaining confidential information is considered the greatest threat to MHA use [65]. Tools are available that can increase device security [50]; however, threats to privacy are continually emerging and endangering data security. For example, identity cannot be confirmed unless a video-calling app is used and personal devices are easily lost or stolen, leaving data vulnerable [50]. Karcher et al [50] determined that the greatest threat to patient privacy in MHA use was the possibility of confidential information being shared with third parties, whether via patient or clinician devices. Hacking of secure devices and new viruses were also identified as challenges to a secure patient database on MHAs. As well as being its own point for consideration when overcoming the challenges of evaluating the clinical effectiveness of MHAs, the complex legal and ethical considerations involved in MHA use are a consideration for clinicians themselves [50]. This has been highlighted in recent news, where contact tracing apps used to help curb the spread of COVID-19 have been the subject of widespread debate.

From a more pragmatic standpoint, MHA cost may be a factor in choosing the right MHA for BD, with 76% of people surveyed reporting interest in using their mobile phones for mental health monitoring and self-management if the MHA was free of charge [15]. Moreover, Larsen et al [66] reported that an MHA clinically relevant for depression is being removed from the market every 2.9 days. This furthers the challenge faced by both patients and clinicians in trying to identify a relevant and appropriate MHA for BD. If the MHA for BD is to be paid for and could be removed from the market without warning, it is difficult to justify its recommendation and purchase.

Limitations

This review had its own limitations. Only 7 MHAs for BD were evaluated, somewhat limiting the generalizability of the results

of this review. Furthermore, only 5 studies commented on the future of the development and evaluation of MHAs for BD.

Conclusion

The studies in our review focused on patient monitoring as an indicator for process and outcome evaluation in MHAs for BD. They based their conclusions on whether the app improved assessment scores rather than interviewing patients on their experience of using the app. Although this is suggested to be a reliable way of measuring the process and outcome values, as modern medicine shifts to holistic patient-centered care, more emphasis should be put on users' experiences rather than quantitative outcomes. In the long term, this will make patients feel respected and involved in the design of MHAs for BD, increasing adherence rates in both the short and long term.

Personalized medicine is a rapidly emerging movement in the field of health care. It is defined as a move away from the "one size fits all" approach to treatment, with new approaches and targeted therapies allowing for flexibility in the management of diseases. With this in mind, more MHAs for BD should be easily available in order to encourage patient choice and freedom to choose an MHA that is best suited to them. At the moment, MONARCA dominates the market, reducing the range and scope of MHAs for BD. As NHS England suggests [60], it can be difficult for an intervention to address both depressive and manic symptoms given the complexity of BD. This is all the more reason to develop a wider variety of apps, with some apps perhaps only focusing on either mania or depression.

The field of MHAs for BD shows promise in both improving patient care and creating a more cost-effective health care service [10]. However, as with any new development in health care, it must be appropriately evaluated and regulated. By encouraging patient co-design and co-evaluation, we can develop a new frontier in personalized digital health, while improving patient experience and care.

Authors' Contributions

IT: acquisition of data, analysis and interpretation of data, drafting the article, and final approval of the version to be published. EC: acquisition of data, analysis and interpretation of data, drafting the article, and final approval of the version to be published. KG: design of the study, revising the manuscript critically for important intellectual content, and final approval of the version to be published. PK: design of the study, acquisition of data, analysis and interpretation of data, revising the manuscript critically for important intellectual content, and final approval of the version to be published. JS: concept of the study, revising the manuscript critically for important intellectual content, and final approval of the version to be published. EBM: concept and design of the study, interpretation of the data, revising the manuscript critically for important intellectual content, and final approval of the version to be published. ANS: concept and design of the study, revising the manuscript critically for important intellectual content, and final approval of the version to be published.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE and PsycINFO searches.

[[DOCX File , 46 KB - jmir_v24i3e29114_app1.docx](#)]

Multimedia Appendix 2

Study characteristics.

[\[DOCX File, 49 KB - jmir_v24i3e29114_app2.docx\]](#)

Multimedia Appendix 3

Study designs and outcomes.

[\[DOCX File, 48 KB - jmir_v24i3e29114_app3.docx\]](#)

References

1. Medici CR, Videbech P, Gustafsson LN, Munk-Jørgensen P. Mortality and secular trend in the incidence of bipolar disorder. *J Affect Disord* 2015 Sep 01;183:39-44. [doi: [10.1016/j.jad.2015.04.032](https://doi.org/10.1016/j.jad.2015.04.032)] [Medline: [26001661](https://pubmed.ncbi.nlm.nih.gov/26001661/)]
2. Drancourt N, Etain B, Lajnef M, Henry C, Raust A, Cochet B, et al. Duration of untreated bipolar disorder: missed opportunities on the long road to optimal treatment. *Acta Psychiatr Scand* 2013 Feb;127(2):136-144. [doi: [10.1111/j.1600-0447.2012.01917.x](https://doi.org/10.1111/j.1600-0447.2012.01917.x)] [Medline: [22901015](https://pubmed.ncbi.nlm.nih.gov/22901015/)]
3. Dagan J, Signorini G, Niessen O, Bani M, Pastore A, Girolamo GD, et al. Meta-analysis of the Interval between the Onset and Management of Bipolar Disorder. *Can J Psychiatry* 2017 Apr;62(4):247-258 [FREE Full text] [doi: [10.1177/0706743716656607](https://doi.org/10.1177/0706743716656607)] [Medline: [27462036](https://pubmed.ncbi.nlm.nih.gov/27462036/)]
4. 65th World Health Assembly. Global burden of mental disorders and the need for a comprehensive, coordinated response from health and social sectors at the country level: report by the Secretariat. World Health Organisation. 2012. URL: <https://apps.who.int/iris/handle/10665/78898> [accessed 2021-03-10]
5. Patel R, Shetty H, Jackson R, Broadbent M, Stewart R, Boydell J, et al. Delays before Diagnosis and Initiation of Treatment in Patients Presenting to Mental Health Services with Bipolar Disorder. *PLoS One* 2015;10(5):e0126530 [FREE Full text] [doi: [10.1371/journal.pone.0126530](https://doi.org/10.1371/journal.pone.0126530)] [Medline: [25992560](https://pubmed.ncbi.nlm.nih.gov/25992560/)]
6. Poushter J. Smartphone Ownership and Internet Usage Continues to Climb in Emerging Economies. Pew Research Center. 2016. URL: <https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/> [accessed 2021-03-10]
7. Children and parents: media use and attitudes report: 2013. Ofcom. 2013. URL: <https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/children-parents-oct-2013> [accessed 2021-03-10]
8. Rosenfeld A, Pendse S, Nugent NR. How mobile health applications can help treat depression. *The Brown University Child and Adolescent Behavior Letter* 2017 Aug 22;33(9):1-6. [doi: [10.1002/cbl.30236](https://doi.org/10.1002/cbl.30236)]
9. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. *Mhealth* 2018;4:6 [FREE Full text] [doi: [10.21037/mhealth.2018.03.02](https://doi.org/10.21037/mhealth.2018.03.02)] [Medline: [29682510](https://pubmed.ncbi.nlm.nih.gov/29682510/)]
10. Ben-Zeev D, Buck B, Hallgren K, Drake RE. Effect of Mobile Health on In-person Service Use Among People With Serious Mental Illness. *Psychiatr Serv* 2019 Jun 01;70(6):507-510. [doi: [10.1176/appi.ps.201800542](https://doi.org/10.1176/appi.ps.201800542)] [Medline: [30947636](https://pubmed.ncbi.nlm.nih.gov/30947636/)]
11. Roberts LW, Chan S, Torous J. New tests, new tools: mobile and connected technologies in advancing psychiatric diagnosis. *NPJ Digit Med* 2018;1:20176 [FREE Full text] [doi: [10.1038/s41746-017-0006-0](https://doi.org/10.1038/s41746-017-0006-0)] [Medline: [31304350](https://pubmed.ncbi.nlm.nih.gov/31304350/)]
12. Lazar MA, Pan Z, Ragguett R, Lee Y, Subramaniapillai M, Mansur RB, et al. Digital revolution in depression: A technologies update for clinicians. *Personalized Medicine in Psychiatry* 2017 Dec;4-6:1-6. [doi: [10.1016/j.pmip.2017.09.001](https://doi.org/10.1016/j.pmip.2017.09.001)]
13. Tal A, Torous J. The digital mental health revolution: Opportunities and risks. *Psychiatr Rehabil J* 2017 Sep;40(3):263-265. [doi: [10.1037/prj0000285](https://doi.org/10.1037/prj0000285)] [Medline: [28891658](https://pubmed.ncbi.nlm.nih.gov/28891658/)]
14. Das Gupta R, Guest JF. Annual cost of bipolar disorder to UK society. *Br J Psychiatry* 2002 Mar;180:227-233. [doi: [10.1192/bjp.180.3.227](https://doi.org/10.1192/bjp.180.3.227)] [Medline: [11872515](https://pubmed.ncbi.nlm.nih.gov/11872515/)]
15. Proudfoot J. The future is in our hands: the role of mobile phones in the prevention and management of mental disorders. *Aust N Z J Psychiatry* 2013 Feb;47(2):111-113. [doi: [10.1177/0004867412471441](https://doi.org/10.1177/0004867412471441)] [Medline: [23382507](https://pubmed.ncbi.nlm.nih.gov/23382507/)]
16. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. *Evid Based Ment Health* 2019 Feb;22(1):4-9 [FREE Full text] [doi: [10.1136/ebmental-2018-300069](https://doi.org/10.1136/ebmental-2018-300069)] [Medline: [30635262](https://pubmed.ncbi.nlm.nih.gov/30635262/)]
17. O'Brien KK, Colquhoun H, Levac D, Baxter L, Tricco AC, Straus S, et al. Advancing scoping study methodology: a web-based survey and consultation of perceptions on terminology, definition and methodological steps. *BMC Health Serv Res* 2016 Jul 26;16:305 [FREE Full text] [doi: [10.1186/s12913-016-1579-z](https://doi.org/10.1186/s12913-016-1579-z)] [Medline: [27461419](https://pubmed.ncbi.nlm.nih.gov/27461419/)]
18. Anderson S, Allen P, Peckham S, Goodwin N. Asking the right questions: scoping studies in the commissioning of research on the organisation and delivery of health services. *Health Res Policy Syst* 2008 Jul 09;6:7 [FREE Full text] [doi: [10.1186/1478-4505-6-7](https://doi.org/10.1186/1478-4505-6-7)] [Medline: [18613961](https://pubmed.ncbi.nlm.nih.gov/18613961/)]
19. 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](https://doi.org/10.1080/1364557032000119616)]
20. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci* 2010 Sep 20;5:69 [FREE Full text] [doi: [10.1186/1748-5908-5-69](https://doi.org/10.1186/1748-5908-5-69)] [Medline: [20854677](https://pubmed.ncbi.nlm.nih.gov/20854677/)]

21. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. *EFI* 2018 Dec 18;34(4):285-291. [doi: [10.3233/efi-180221](https://doi.org/10.3233/efi-180221)]
22. Hidalgo-Mazzei D, Mateu A, Reinares M, Murru A, Del Mar Bonnín C, Varo C, et al. Psychoeducation in bipolar disorder with a SIMPLE smartphone application: Feasibility, acceptability and satisfaction. *J Affect Disord* 2016 Aug;200:58-66. [doi: [10.1016/j.jad.2016.04.042](https://doi.org/10.1016/j.jad.2016.04.042)] [Medline: [27128358](https://pubmed.ncbi.nlm.nih.gov/27128358/)]
23. Bardram J, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing LV. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. In: CHI '13: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 2, 2013; Paris, France p. 2627-2636. [doi: [10.1145/2470654.2481364](https://doi.org/10.1145/2470654.2481364)]
24. Schärer LO, Krienke UJ, Graf S, Meltzer K, Langosch JM. Validation of life-charts documented with the personal life-chart app - a self-monitoring tool for bipolar disorder. *BMC Psychiatry* 2015 Mar 14;15:49 [FREE Full text] [doi: [10.1186/s12888-015-0414-0](https://doi.org/10.1186/s12888-015-0414-0)] [Medline: [25885225](https://pubmed.ncbi.nlm.nih.gov/25885225/)]
25. Osmani V, Maxhuni A, Grünerbl A, Lukowicz P, Haring C, Mayora O. Monitoring activity of patients with bipolar disorder using smart phones. In: MoMM '13: Proceedings of International Conference on Advances in Mobile Computing & Multimedia. 2013 Presented at: International Conference on Advances in Mobile Computing & Multimedia; December 2-4, 2013; Vienna, Austria p. 85-92. [doi: [10.1145/2536853.2536882](https://doi.org/10.1145/2536853.2536882)]
26. Schwartz S, Schultz S, Reider A, Saunders EFH. Daily mood monitoring of symptoms using smartphones in bipolar disorder: A pilot study assessing the feasibility of ecological momentary assessment. *J Affect Disord* 2016 Feb;191:88-93 [FREE Full text] [doi: [10.1016/j.jad.2015.11.013](https://doi.org/10.1016/j.jad.2015.11.013)] [Medline: [26655117](https://pubmed.ncbi.nlm.nih.gov/26655117/)]
27. Faurholt-Jepsen M, Frost M, Ritz C, Christensen EM, Jacoby AS, Mikkelsen RL, et al. Daily electronic self-monitoring in bipolar disorder using smartphones - the MONARCA I trial: a randomized, placebo-controlled, single-blind, parallel group trial. *Psychol Med* 2015 Oct;45(13):2691-2704. [doi: [10.1017/S0033291715000410](https://doi.org/10.1017/S0033291715000410)] [Medline: [26220802](https://pubmed.ncbi.nlm.nih.gov/26220802/)]
28. Tsanas A, Saunders KEA, Bilderbeck AC, Palmius N, Osipov M, Clifford GD, Goodwin, et al. Daily longitudinal self-monitoring of mood variability in bipolar disorder and borderline personality disorder. *J Affect Disord* 2016 Nov 15;205:225-233 [FREE Full text] [doi: [10.1016/j.jad.2016.06.065](https://doi.org/10.1016/j.jad.2016.06.065)] [Medline: [27449555](https://pubmed.ncbi.nlm.nih.gov/27449555/)]
29. Faurholt-Jepsen M, Ritz C, Frost M, Mikkelsen RL, Margrethe Christensen E, Bardram J, et al. Mood instability in bipolar disorder type I versus type II-continuous daily electronic self-monitoring of illness activity using smartphones. *J Affect Disord* 2015 Nov 01;186:342-349. [doi: [10.1016/j.jad.2015.06.026](https://doi.org/10.1016/j.jad.2015.06.026)] [Medline: [26277270](https://pubmed.ncbi.nlm.nih.gov/26277270/)]
30. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as an electronic biomarker of illness activity in bipolar disorder. *Bipolar Disord* 2015 Nov;17(7):715-728. [doi: [10.1111/bdi.12332](https://doi.org/10.1111/bdi.12332)] [Medline: [26395972](https://pubmed.ncbi.nlm.nih.gov/26395972/)]
31. Guidi A, Salvi S, Ottaviano M, Gentili C, Bertschy G, de Rossi D, et al. Smartphone Application for the Analysis of Prosodic Features in Running Speech with a Focus on Bipolar Disorders: System Performance Evaluation and Case Study. *Sensors (Basel)* 2015 Nov 06;15(11):28070-28087 [FREE Full text] [doi: [10.3390/s151128070](https://doi.org/10.3390/s151128070)] [Medline: [26561811](https://pubmed.ncbi.nlm.nih.gov/26561811/)]
32. Faurholt-Jepsen M, Frost M, Vinberg M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as objective measures of bipolar disorder symptoms. *Psychiatry Res* 2014 Jun 30;217(1-2):124-127. [doi: [10.1016/j.psychres.2014.03.009](https://doi.org/10.1016/j.psychres.2014.03.009)] [Medline: [24679993](https://pubmed.ncbi.nlm.nih.gov/24679993/)]
33. Beiwinkel T, Kindermann S, Maier A, Kerl C, Mook J, Barbian G, et al. Using Smartphones to Monitor Bipolar Disorder Symptoms: A Pilot Study. *JMIR Ment Health* 2016 Jan 06;3(1):e2 [FREE Full text] [doi: [10.2196/mental.4560](https://doi.org/10.2196/mental.4560)] [Medline: [26740354](https://pubmed.ncbi.nlm.nih.gov/26740354/)]
34. Hidalgo-Mazzei D, Mateu A, Reinares M, Matic A, Vieta E, Colom F. Internet-based psychological interventions for bipolar disorder: Review of the present and insights into the future. *J Affect Disord* 2015 Dec 01;188:1-13. [doi: [10.1016/j.jad.2015.08.005](https://doi.org/10.1016/j.jad.2015.08.005)] [Medline: [26342885](https://pubmed.ncbi.nlm.nih.gov/26342885/)]
35. Young RC, Biggs JT, Ziegler VE, Meyer DA. A rating scale for mania: reliability, validity and sensitivity. *Br J Psychiatry* 1978 Nov;133:429-435. [doi: [10.1192/bjp.133.5.429](https://doi.org/10.1192/bjp.133.5.429)] [Medline: [728692](https://pubmed.ncbi.nlm.nih.gov/728692/)]
36. 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](https://doi.org/10.1136/jnnp.23.1.56)] [Medline: [14399272](https://pubmed.ncbi.nlm.nih.gov/14399272/)]
37. Altman EG, Hedeker D, Peterson JL, Davis JM. The Altman Self-Rating Mania Scale. *Biological Psychiatry* 1997 Nov 15;42(10):948-955. [doi: [10.1016/S0006-3223\(96\)00548-3](https://doi.org/10.1016/S0006-3223(96)00548-3)] [Medline: [9359982](https://pubmed.ncbi.nlm.nih.gov/9359982/)]
38. Rush AJ, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 2003 Sep 01;54(5):573-583. [doi: [10.1016/s0006-3223\(02\)01866-8](https://doi.org/10.1016/s0006-3223(02)01866-8)] [Medline: [12946886](https://pubmed.ncbi.nlm.nih.gov/12946886/)]
39. Madera J, Such P, Zhang P, Baker RA, Grande I. Use of the Functioning Assessment Short Test (FAST) in defining functional recovery in bipolar I disorder. Post-hoc analyses of long-term studies of aripiprazole once monthly as maintenance treatment. *Neuropsychiatr Dis Treat* 2019;15:2325-2338 [FREE Full text] [doi: [10.2147/NDT.S209700](https://doi.org/10.2147/NDT.S209700)] [Medline: [31616148](https://pubmed.ncbi.nlm.nih.gov/31616148/)]
40. Endler NS, Parker JDA. Assessment of multidimensional coping: Task, emotion, and avoidance strategies. *Psychological Assessment* 1994 Mar;6(1):50-60. [doi: [10.1037/1040-3590.6.1.50](https://doi.org/10.1037/1040-3590.6.1.50)]

41. 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](https://doi.org/10.1001/archinte.166.10.1092)] [Medline: [16717171](https://pubmed.ncbi.nlm.nih.gov/16717171/)]
42. Rush AJ, Carmody T, Reimnitz P. The Inventory of Depressive Symptomatology (IDS): Clinician (IDS-C) and Self-Report (IDS-SR) ratings of depressive symptoms. *Int. J. Method. Psychiat. Res* 2006 Jun;9(2):45-59. [doi: [10.1002/mpr.79](https://doi.org/10.1002/mpr.79)]
43. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. *J Pers Soc Psychol* 2003 Apr;84(4):822-848. [doi: [10.1037/0022-3514.84.4.822](https://doi.org/10.1037/0022-3514.84.4.822)] [Medline: [12703651](https://pubmed.ncbi.nlm.nih.gov/12703651/)]
44. Olsen LR, Jensen DV, Noerholm V, Martiny K, Bech P. The internal and external validity of the Major Depression Inventory in measuring severity of depressive states. *Psychol Med* 2003 Feb;33(2):351-356. [doi: [10.1017/s0033291702006724](https://doi.org/10.1017/s0033291702006724)] [Medline: [12622314](https://pubmed.ncbi.nlm.nih.gov/12622314/)]
45. Tan X, Patel I, Chang J. Review of the four item Morisky Medication Adherence Scale (MMAS-4) and eight item Morisky Medication Adherence Scale (MMAS-8). *Innov Pharm* 2014 Jan 01;5(3):347. [doi: [10.24926/iip.v5i3.347](https://doi.org/10.24926/iip.v5i3.347)]
46. Cohen S. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. *The social psychology of health*. Thousand Oaks, CA: Sage Publications, Inc; 1988:31-67.
47. No authors listed. The World Health Organization Quality of Life assessment (WHOQOL): position paper from the World Health Organization. *Soc Sci Med* 1995 Nov;41(10):1403-1409. [doi: [10.1016/0277-9536\(95\)00112-k](https://doi.org/10.1016/0277-9536(95)00112-k)] [Medline: [8560308](https://pubmed.ncbi.nlm.nih.gov/8560308/)]
48. Streicher A, Smeddinck J. Personalized and Adaptive Serious Games. In: Dörner R, Göbel S, Kickmeier-Rust M, Masuch M, Zweig K, editors. *Entertainment Computing and Serious Games*. Lecture Notes in Computer Science, vol 9970. Cham: Springer; 2016:332-377.
49. Smeddinck J, Herrlich M, Malaka R. Exergames for Physiotherapy and Rehabilitation: A Medium-term Situated Study of Motivational Aspects and Impact on Functional Reach. In: CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: 33rd Annual ACM Conference on Human Factors in Computing Systems; December 2-4, 2013; Seoul, Republic of Korea p. 4143-4146. [doi: [10.1145/2702123.2702598](https://doi.org/10.1145/2702123.2702598)]
50. Karcher NR, Presser NR. Ethical and Legal Issues Addressing the Use of Mobile Health (mHealth) as an Adjunct to Psychotherapy. *Ethics & Behavior* 2016 Sep 03;28(1):1-22. [doi: [10.1080/10508422.2016.1229187](https://doi.org/10.1080/10508422.2016.1229187)]
51. New simpler and faster assessment process for digital health technologies launched for the NHS and social care. NHS. 2021. URL: <https://www.nhsx.nhs.uk/news/new-simpler-and-faster-assessment-process-for-digital-health-technologies-launched-for-the-nhs-and-social-care/> [accessed 2021-06-24]
52. How to use the DTAC. NHS. URL: <https://www.nhsx.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac/how-to-use-the-dtac/> [accessed 2021-06-24]
53. NHS Apps Library. NHS. URL: <https://www.nhs.uk/apps-library/> [accessed 2021-06-24]
54. Byambasuren O, Beller E, Glasziou P. Current Knowledge and Adoption of Mobile Health Apps Among Australian General Practitioners: Survey Study. *JMIR Mhealth Uhealth* 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: [10.2196/13199](https://doi.org/10.2196/13199)] [Medline: [31199343](https://pubmed.ncbi.nlm.nih.gov/31199343/)]
55. O'Connell C. 23% of Users Abandon an App After One Use. DZone. 2016. URL: <https://dzone.com/articles/23-of-users-abandon-an-app-after-one-use> [accessed 2021-03-01]
56. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The Use and Effectiveness of Mobile Apps for Depression: Results From a Fully Remote Clinical Trial. *J Med Internet Res* 2016 Dec 20;18(12):e330 [FREE Full text] [doi: [10.2196/jmir.6482](https://doi.org/10.2196/jmir.6482)] [Medline: [27998876](https://pubmed.ncbi.nlm.nih.gov/27998876/)]
57. Kenny R, Dooley B, Fitzgerald A. Ecological Momentary Assessment of Adolescent Problems, Coping Efficacy, and Mood States Using a Mobile Phone App: An Exploratory Study. *JMIR Ment Health* 2016 Nov 29;3(4):e51 [FREE Full text] [doi: [10.2196/mental.6361](https://doi.org/10.2196/mental.6361)] [Medline: [27899340](https://pubmed.ncbi.nlm.nih.gov/27899340/)]
58. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. *Evid Based Ment Health* 2018 Aug 05;21(3):116-119. [doi: [10.1136/eb-2018-102891](https://doi.org/10.1136/eb-2018-102891)] [Medline: [29871870](https://pubmed.ncbi.nlm.nih.gov/29871870/)]
59. Nicholas J, Boydell K, Christensen H. mHealth in psychiatry: time for methodological change. *Evid Based Ment Health* 2016 May;19(2):33-34. [doi: [10.1136/eb-2015-102278](https://doi.org/10.1136/eb-2015-102278)] [Medline: [27044849](https://pubmed.ncbi.nlm.nih.gov/27044849/)]
60. Improving Outcomes Through Personalised Medicine. NHS. URL: <https://www.england.nhs.uk/wp-content/uploads/2016/09/improving-outcomes-personalised-medicine.pdf> [accessed 2021-06-24]
61. Blankenhagel KJ. Identifying Usability Challenges of eHealth Applications for People with Mental Disorders: Errors and Design Recommendations. In: PervasiveHealth'19: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2019 Presented at: 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2019; Trento, Italy p. 91-100. [doi: [10.1145/3329189.3329195](https://doi.org/10.1145/3329189.3329195)]
62. Exploring the UK's digital divide. Office for National Statistics. 2019. URL: <https://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/articles/exploringtheuksdigitaldivide/2019-03-04> [accessed 2021-06-04]
63. Ennis L, Rose D, Denis M, Pandit N, Wykes T. Can't surf, won't surf: the digital divide in mental health. *J Ment Health* 2012 Aug 19;21(4):395-403 [FREE Full text] [doi: [10.3109/09638237.2012.689437](https://doi.org/10.3109/09638237.2012.689437)] [Medline: [22712756](https://pubmed.ncbi.nlm.nih.gov/22712756/)]

64. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AAR, et al. Finding a depression app: a review and content analysis of the depression app marketplace. *JMIR Mhealth Uhealth* 2015 Feb 16;3(1):e16 [FREE Full text] [doi: [10.2196/mhealth.3713](https://doi.org/10.2196/mhealth.3713)] [Medline: [25689790](https://pubmed.ncbi.nlm.nih.gov/25689790/)]
65. Guidelines for the Practice of Telepsychology. American Psychological Association. URL: <https://www.apa.org/practice/guidelines/telepsychology> [accessed 2021-03-01]
66. Larsen ME, Nicholas J, Christensen H. Quantifying App Store Dynamics: Longitudinal Tracking of Mental Health Apps. *JMIR Mhealth Uhealth* 2016 Aug 09;4(3):e96 [FREE Full text] [doi: [10.2196/mhealth.6020](https://doi.org/10.2196/mhealth.6020)] [Medline: [27507641](https://pubmed.ncbi.nlm.nih.gov/27507641/)]

Abbreviations

BD: bipolar disorder

DTAC: Digital Technology Assessment Criteria

MHA: mental health app

MMAT: Mixed Methods Appraisal Tool

NHS: National Health Service

Edited by A Mavragani; submitted 26.03.21; peer-reviewed by N Khalili-Mahani, F Lobban, H Mehdizadeh; comments to author 07.05.21; revised version received 28.07.21; accepted 01.12.21; published 23.03.22.

Please cite as:

Tatham I, Clarke E, Grieve KA, Kaushal P, Smeddinck J, Millar EB, Sharma AN

Process and Outcome Evaluations of Smartphone Apps for Bipolar Disorder: Scoping Review

J Med Internet Res 2022;24(3):e29114

URL: <https://www.jmir.org/2022/3/e29114>

doi: [10.2196/29114](https://doi.org/10.2196/29114)

PMID: [35319470](https://pubmed.ncbi.nlm.nih.gov/35319470/)

©Iona Tatham, Ellisiv Clarke, Kelly Ann Grieve, Pulkit Kaushal, Jan Smeddinck, Evelyn Barron Millar, Aditya Narain Sharma. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 23.03.2022. 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.

Review

Digital Assessment Tools Using Animation Features to Quantify Alcohol Consumption: Systematic App Store and Literature Review

Veronika Wiemker¹; Maria Neufeld², MSc; Anna Bunova³, MD; Ina Danquah¹, PhD; Carina Ferreira-Borges², PhD; Stefan Konigorski^{4,5}, PhD; Ankit Rastogi¹; Charlotte Probst^{1,6,7}, PhD

¹Heidelberg Institute of Global Health, Medical Faculty and University Hospital, Heidelberg University, Heidelberg, Germany

²WHO European Office for the Prevention and Control of Noncommunicable Diseases, Moscow, Russian Federation

³National Medical Research Center for Therapy and Preventive Medicine of the Ministry of Health of the Russian Federation, Moscow, Russian Federation

⁴Digital Health Center, Hasso Plattner Institute for Digital Engineering, University of Potsdam, Potsdam, Germany

⁵Hasso Plattner Institute for Digital Health at Mount Sinai, Icahn School of Medicine at Mount Sinai, New York, NY, United States

⁶Institute for Mental Health Policy Research, Centre for Addiction and Mental Health, Toronto, ON, Canada

⁷Department of Psychiatry, University of Toronto, Toronto, ON, Canada

Corresponding Author:

Charlotte Probst, PhD

Heidelberg Institute of Global Health

Medical Faculty and University Hospital

Heidelberg University

Im Neuenheimer Feld 130.3

Heidelberg, 69120

Germany

Phone: 49 6221 56 5040

Fax: 49 6221 56 5948

Email: charlotte.probst@uni-heidelberg.de

Abstract

Background: Accurate and user-friendly assessment tools for quantifying alcohol consumption are a prerequisite for effective interventions to reduce alcohol-related harm. Digital assessment tools (DATs) that allow the description of consumed alcoholic drinks through animation features may facilitate more accurate reporting than conventional approaches.

Objective: This review aims to identify and characterize freely available DATs in English or Russian that use animation features to support the quantitative assessment of alcohol consumption (alcohol DATs) and determine the extent to which such tools have been scientifically evaluated in terms of feasibility, acceptability, and validity.

Methods: Systematic English and Russian searches were conducted in iOS and Android app stores and via the Google search engine. Information on the background and content of eligible DATs was obtained from app store descriptions, websites, and test completions. A systematic literature review was conducted in Embase, MEDLINE, PsycINFO, and Web of Science to identify English-language studies reporting the feasibility, acceptability, and validity of animation-using alcohol DATs. Where possible, the evaluated DATs were accessed and assessed. Owing to the high heterogeneity of study designs, results were synthesized narratively.

Results: We identified 22 eligible alcohol DATs in English, 3 (14%) of which were also available in Russian. More than 95% (21/22) of tools allowed the choice of a beverage type from a visually displayed selection. In addition, 36% (8/22) of tools enabled the choice of a drinking vessel. Only 9% (2/22) of tools allowed the simulated interactive pouring of a drink. For none of the tools published evaluation studies were identified in the literature review. The systematic literature review identified 5 exploratory studies evaluating the feasibility, acceptability, and validity of 4 animation-using alcohol DATs, 1 (25%) of which was available in the searched app stores. The evaluated tools reached moderate to high scores on user rating scales and showed fair to high convergent validity when compared with established assessment methods.

Conclusions: Animation-using alcohol DATs are available in app stores and on the web. However, they often use nondynamic features and lack scientific background information. Explorative study data suggest that such tools might enable the user-friendly

and valid assessment of alcohol consumption and could thus serve as a building block in the reduction of alcohol-attributable health burden worldwide.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020172825; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020172825

(*J Med Internet Res* 2022;24(3):e28927) doi:[10.2196/28927](https://doi.org/10.2196/28927)

KEYWORDS

alcohol consumption; harmful and hazardous drinking; screening; assessment methods; eHealth; mobile apps; visualization; animation features; AUDIT; primary health care

Introduction

Background

Alcohol-related injuries and diseases are major causes of morbidity and mortality worldwide, although, at least in theory, they are fully preventable [1]. The well-directed implementation of monitoring, prevention, and treatment programs requires accurate assessment tools to quantify the users' alcohol consumption. To date, consumption assessments are generally based on standardized self-report questionnaires or brief interviews. At the population level, they form the foundation for public health monitoring, quantification of alcohol-attributable harm, and evaluation of alcohol policies. At the individual level, they constitute the cornerstone of effective harm reduction strategies such as screening and brief intervention (SBI) programs. SBI programs link the routine administration of a screening tool to identify harmful or hazardous drinking, often a questionnaire, to a tailored brief intervention, most commonly comprising a short motivational interview or structured advice [2]. They have been shown to be highly effective in reducing excessive drinking among adults [3,4] and are recommended in national and international policy guidelines for reducing alcohol-attributable harm [5,6]. However, the implementation of SBI programs in public health systems remains low [7,8]. Relevant implementation barriers include a perceived lack of knowledge or skills among health care professionals and environmental context factors such as time restrictions and limited resources [9].

Although the measures used in epidemiological surveys differ between countries and regions [10-12], to date, most of them ultimately require the counting of *standard drinks* consumed. Routine screening tools for primary care such as the Alcohol Use Disorders Identification Test (AUDIT) developed by the World Health Organization [13] and its abbreviated form AUDIT-Consumption (AUDIT-C) [14] also rely on this concept. The standard drink, defined as a beverage volume containing a fixed amount of pure alcohol, facilitates the comparison and assessment of alcohol quantities across different beverage types with varying alcohol content. However, this concept is problematic for 2 main reasons. First, standard drink sizes differ considerably between countries, cultures, and settings, with national definitions varying even within Europe from 8 g of pure alcohol in the United Kingdom to 20 g in Austria [15]. In fact, the majority of countries worldwide do not have an official definition [15]. In addition, relevant AUDIT items are often not adapted to account for differing national standard drink sizes, as required in the AUDIT manual [13,16].

Second, even in countries where the standard drink concept is officially used to standardize the size of retail alcohol, consumers are often not acquainted with the concept and many are unable to convert their consumption correctly [17,18]. For instance, when asked to pour their usual drink and subsequently estimate the number of standard drinks it contained, primary health care patients in the United Kingdom over- or underestimated their actual drink size by at least 0.5 standard drinks in more than half of the cases [19]. In a study conducted among health care professionals in the United States, <20% of the interviewed clinicians could correctly state the volume of a standard drink of liquor [20]. Alongside other known biasing factors, such as memory and social desirability bias or underreporting because of alcohol-related stigma, this might contribute to the considerable underestimation of the total alcohol recorded through official statistics by approximately 50% in nationally representative surveys [21-24].

Evidence suggests that the assessment results of digital and traditional administration modes are comparable in epidemiological surveys as well as in screening situations [25-27]. Promises of digitally administered tools, such as increased standardization and time efficiency, adaptability of the assessment flow based on user input, and seamless integration with electronic health records [28], may thus help address central SBI implementation barriers [29]. Importantly, digital assessment tools (DATs) can replace the standard drink concept by using individualized, interactive animation features to assess the type and amount of alcohol consumed. Related research fields such as nutrition epidemiology have already recognized the usefulness of visualization features to improve the quantification of consumption [30,31].

Currently, there is a growing body of literature focusing on the effectiveness and availability of evidence-based alcohol reduction apps [32,33]. These apps often contain a screening part quantifying the user's consumption, which might be text based [34] or based on interactive animations [35]. However, to the best of our knowledge, the current availability of interactive animation features in alcohol DATs has not been systematically evaluated. There is also no systematic review of the effects of such features on assessment feasibility, acceptability, and validity.

Research Questions

This review seeks to answer the following two questions with a focus on DATs quantifying alcohol consumption (alcohol DATs), which use animation to support users in describing their consumption:

1. What freely available animation-using alcohol DATs exist in the English or Russian language, and what are their core characteristics?
2. To what extent have such tools been scientifically evaluated in terms of feasibility, acceptability, and validity?

Methods

Study Design

This systematic review was performed in 2 parts. In part 1, an app store search and a web-based search were conducted to identify existing freely available alcohol DATs. Part 2 comprised a systematic literature search to identify studies that evaluated the feasibility, acceptability, and validity of animation-using alcohol DATs.

The study protocol was published in PROSPERO (International Prospective Register of Systematic Reviews; registration number: CRD42020172825) [36]. We adhered to the standards set out in PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2020 [37]. Where applicable, we also followed the recommendations for methodological reporting of systematic searches in app store environments proposed by Grainger et al [38].

Part 1: Existing DATs

Eligibility Criteria

This review focused on animation-using alcohol DATs, defined as tools that allow the assessment and quantification of the user's alcohol consumption via an electronic display device (a PC, laptop, or a mobile device). Aiming to include any alcohol DATs using interactive visualizations as opposed to purely text-based quantification tools, a broad definition of the term *animation* was chosen. Specifically, tools were considered to be using animations if they included ≥ 1 of the following features: (1) selection of a drink or a beverage type from a number of visually displayed options; (2) selection of a drinking vessel from a number of visually displayed glasses and, in some instances, bottles; and (3) simulated interactive *pouring* of a drink—that is, continuously adjusting the beverage level displayed in the chosen drinking vessel. The availability of each of the listed features was recorded to classify the complexity of the animation used. In addition, tools had to allow for the quantification of the user's alcohol consumption over a defined reference period or occasion, be available in English or Russian, and be accessible free of charge. English-language apps form the largest language group among all apps available in the iOS App Store and the Google Play Store [39]. The mentioned app stores represent approximately 95% of the app market share worldwide [40] and offer $\geq 95\%$ of the available apps free of charge [41,42]. Given the language background of the authors, the review additionally focuses on Russian-language apps. Russia has one of the highest proportions of alcohol-attributable mortality worldwide, and digital health interventions might become part of the promising prevention efforts currently taken and underway [43]. Our search for Russian-language alcohol DATs aims to identify relevant Russian-language alcohol DATs and gauge the potential of repeating the systematic search in additional languages in the future.

Search Strategy

The German app store versions of Google Play Store and iOS App Store were searched in June and July 2020, with English as the preferred app language. As app store search functions do not allow the systematic combination of search terms, 4 independent searches were performed on each platform, using the search terms *alcohol*, *alcohol screening*, *alcohol test*, and *drinking*. We recorded the first 250 results per platform using the search term *alcohol*. Given the high overlap between search results and decreasing relevance after the first 50 to 70 results, a maximum of 100 search results were screened for each of the other search terms. The Google search engine was searched in August 2020 using three sets of search terms (*alcohol screening online*, *alcohol test online*, and *drinking test online*). A total of 90 websites were included in the screening. To further explore the extent of regional adaptation in alcohol DATs and potential content differences between national app stores, we conducted additional searches in January and February 2021 in the Russian version of the Google Play and iOS App Store, with Russian as the preferred app language, using translated search terms. All search results were screened. The Russian Google search engine was searched in February 2021.

Screening and Selection of Tools

The URLs and titles of all identified app store entries or websites were recorded. After removing duplicates with identical URLs, the remaining app store descriptions and websites were screened for eligibility. A random sample of 25 English app store entries was independently screened by a second reviewer, and agreement was quantified to ensure the objectivity of the eligibility criteria. After screening, potentially eligible mobile tools were downloaded and completed on mobile devices (for English searches: Huawei Honor 9 Lite LLD-L31, Android version 9 and iPhone SE (2016), iOS version 14.0.1; for Russian searches: Samsung Galaxy Tab A 7.0 SM-T285 8 Gb, Android version 9 and Apple iPad (2018), iOS version 11.2). Web-based tools were completed on the web via the Safari and Google Chrome browsers to determine eligibility.

Data Extraction

The following data were extracted from the app store entries and linked websites and through testing the apps or web-based tools: general information (tool name, developer, responsible organization, link to website, scientific background or development process, country of publisher, year of the last update, and number of downloads), content features (reference period, underlying questionnaire, feedback on the user's consumption quantity, use of standard drink concept, target group, and characteristic additional features), and animation features (availability of abovementioned features and options to adjust further drink characteristics).

Part 2: Studies on Feasibility, Acceptability, and Validity

Eligibility Criteria

Part 2 of the review aimed to identify (1) validation studies comparing animation-using alcohol DATs with established assessment methods (eg, paper–pencil, interview, or web-based

questionnaires such as the AUDIT [13], AUDIT-C [14], Alcohol, Smoking, and Substance Involvement Screening Test [44], Alcohol Timeline Followback [45]; drinking diaries; standardized clinical interviews; or alcohol biomarkers) and (2) studies reporting on feasibility or acceptability of animation-using alcohol DATs. Eligibility was restricted to completed and fully reported studies. The same eligibility criteria for animation-using alcohol DATs were used as in part 1. When it was not possible to determine whether the eligibility criteria for using animations were met or when a study used a sample of participants aged <15 years, the study was excluded. Studies conducted among general and specialized populations, such as patient populations, were eligible. No geographical, language, or time restrictions were applied.

Search Strategy

A systematic literature search was performed using Embase, MEDLINE, PsycINFO, and Web of Science. Search terms (Multimedia Appendix 1) were adapted to the requirements of each web-based database with regard to medical subject headings and wildcards. The searches covered publications from January 2000 to August 2020.

Screening and Selection of Studies

After removing duplicates, titles and abstracts were screened by a first reviewer, and a subsample of 100 records was independently screened by a second reviewer. In a second step, the full texts were obtained to decide about final inclusion.

Data Extraction

Information on general study characteristics (title, authors, year of publication, and type of study), study methods (setting,

design, comparator, sample and recruitment strategy, period of data collection, outcomes, and outcome measurement), main findings, and information on the tested alcohol DATs were extracted. Owing to the high heterogeneity of the study designs, no standardized risk of bias assessment was performed.

Where possible, evaluated DATs were accessed and assessed against the same criteria as the DAT identified in the systematic app store search.

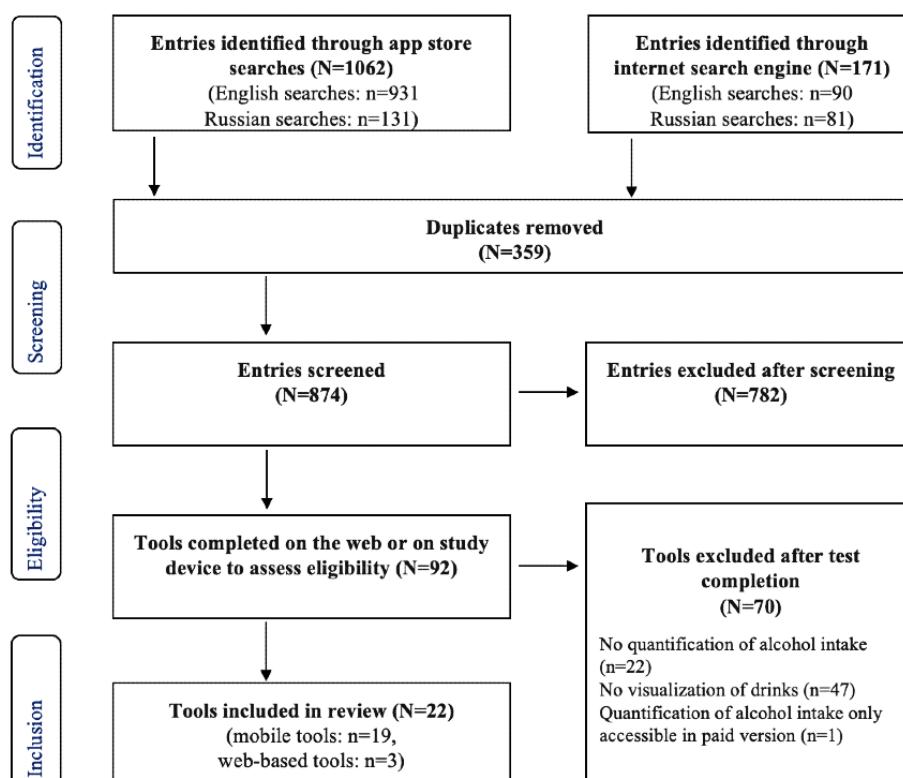
Results

Part 1: Existing DATs—Tools Identified and Included

Overview

A total of 1062 app store entries and 171 web entries were identified through app store and web searches (Figure 1). The searches in the Russian language yielded a much lower number of results than the English-language searches. After removing duplicates, of the 1233 total entries, 874 (70.88%) entries were screened for eligible alcohol DATs. Agreement between the reviewers was 92% for exclusion decisions after screening. A total of 54 mobile tools and 38 web-based tools were considered and tested for final inclusion. Finally, 35% (19/54) of mobile tools and 8% (3/38) of web-based tools were eligible. Of the 19 included tools, 16 (84%) were available in English only, 3 (16%) were available in both English and Russian without adaptations in content [46-48], and none were available in Russian only. All included mobile tools were available in German app stores; all but 16% (3/19) of mobile tools [49-51] could also be downloaded from Russian app stores.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of tool selection.



Content Characteristics

The core characteristics of the 22 alcohol DATs are summarized in Table 1 (detailed in Multimedia Appendix 2) [46-65]. Of the 19 mobile tools, 9 (47%) were available for both iOS and Android operating systems, 5 (26%) were published only in the iOS app store, and 5 (26%) were published only in the Google Play Store. The year of the last update ranged from 2014 to 2020, with 55% (12/22) of tools updated in 2020 or the previous year. The download numbers (only available for Android apps) ranged from ≥ 100 [49] to 50,000 [48,52]. Approximately 36% (8/22) of tools were developed in the United Kingdom [51-57], and 5% (1/22) each in Ireland [49], Canada [58], Russia [47], France [48], Denmark [46], Germany [59], and Japan [50]. The country of origin could not be identified for 18% (4/22) of tools [60-62,66]. There was no information available regarding any regional or cultural adaptation, and for tools available in both Russian and English, no cultural adaptations were evident. Publishing institutions included public actors [51,53,58], registered charities in the field of alcohol use prevention and general health [52,56,63], and private companies [46-48,54,55,59]. No information about the legal status of the publishing institution could be identified for 32% (7/22) of tools [49,50,57,61,62,66]. For only 14% (3/22) of tools, a scientific background and development process was mentioned [51,58,59]. With the exception of 5% (1/22) of tools designed for health care professionals [51], all tools targeted the general adult population, with a focus on individuals wanting to monitor or cut down their alcohol consumption.

Out of 22 identified tools, 3 (14%) were primarily designed to assess risky drinking in a one-time screening [51,53,54] and led to a structured feedback section, including or enabling (1) an estimate of the user's alcohol-related health risk, (2) a comparison of the individual consumption to a relevant guideline or reference group, and (3) additional information on the standard drink concept and alcohol-related health risks. A total of 14% (3/22) of the identified tools were designed to deliver individualized programs to reduce or quit drinking [46,58,59]

and started with a brief prospective [59] or retrospective [46,58] assessment of the user's baseline consumption, followed by a tailored reduction scheme. All remaining tools relied on real-time assessment and were designed to either estimate the users' blood alcohol concentration (*blood alcohol concentration calculators*; 4/22, 18% of tools) [50,60,63,66], count the number of standard drinks at a drinking occasion (*standard drink counters*; 1/22, 5% of tools) [48], or keep track of the alcohol consumed over a longer period (*drinking diaries*; 8/22, 36% of tools) [47,49,52,56,57,61,62]

Although none of the tools relied on the standard drink concept in the assessment part, most (13/22, 59% of tools) referred to this concept in their results and feedback sections [48,51-58,61,63,66,67]. In addition to alcohol-related health risk and consumed alcohol quantity, 45% (10/22) of tools reported money spent on alcohol, calories consumed, and hypothetical money or calories saved by cutting down drinking [48,50,52-56,58,61,63]

Similar to the mobile tools, all 3 included web-based alcohol DATs addressed the general adult population. They were provided by nonprofit organizations from Ireland [64,65] and the United Kingdom [56], with copyright claims absent [56] or dating to the current (2020) [65] or past year [64]. All organizations provided contact details of support services helping to cut down drinking. Out of 22 identified tools, 2 (9%) [56,64] were digital versions of the World Health Organization's AUDIT. The functionality of standard drink calculation was directly embedded into AUDIT-C item 3 ("How many units of alcohol do you drink on a typical day when you are drinking?"), preserving the questionnaire's original item structure. Both led to a detailed feedback section, including AUDIT score and risk category, information on standard drinks, and calories consumed on a typical day. The third tool converted the user's reported consumption into standard drinks [65] and provided additional feedback items, including a comparison of the user's alcohol consumption to a weekly low-risk drinking guideline [68].

Table 1. DATs^a quantifying alcohol consumption (alcohol DATs) in the English language: core characteristics of the included tools (N=22).

Tool name (year of last update ^b ; country)	Animation features			Adjust drinks ^c (n=13)	User feedback		Extra features
	Drinks ^d (n=21)	Vessels ^e (n=9)	Pour ^f (n=2)		Unit of consumption ^g	Additional feedback ^h	
Mobile app: 1-time assessment of risky drinking							
Drinks Meter (2020; United Kingdom) [54]	✓	✓		✓	Standard drinks	Physiology or nutrition	Text-based AUDIT ⁱ ; <i>drink pourer</i> tool
Know Your Numbers (2017; United Kingdom) [51]	✓	✓			Standard drinks	— ^j	Alcohol unit guide
Know Your Units (2017; United Kingdom) [53]	✓			✓	Standard drinks	Physiology or nutrition	Beverage-specific sound animations
Mobile app: individualized program to reduce or quit drinking							
MeSelfControl (2016; Germany) [59]	✓	✓		✓	Alcohol quantity	—	—
ReduceYour Drinking (2015; Denmark) [46]	✓				Alcohol quantity	—	Text-based DATs; available in Russian
Saying When (2016; Canada) [58]	✓	✓	✓		Standard drinks	Positive effect	Explanation of standard drink concept
Mobile app: BAC^k calculator							
alcCalc (2014; Japan) [50]	✓				Alcohol quantity	Physiology or nutrition	—
Alcohol Diary (2019; not provided) [67]	✓				Standard drinks	—	—
Alcohol meter (2019; not provided) [60]			✓	✓	Alcohol quantity	Physiology or nutrition	—
DrinkWatch Unit Checker (2016; United Kingdom) [63]	✓		✓	✓	Standard drinks	Physiology or nutrition; negative effect	—
Mobile app: drinking diary							
AlcoExpert (2019; Russia) [47]	✓			✓	Alcohol quantity	Physiology or nutrition; negative effect	Photorealistic drink images; available in Russian
Alcofy (2020; not provided) [62]	✓	✓		✓	Alcohol quantity	Physiology or nutrition	—
DrinkCoach (2020; United Kingdom) [56]	✓			✓	Standard drinks	Physiology or nutrition; positive effect	Visualized drinking scene; link to animation-enhanced AUDIT
DrinkControl (2020; not provided) [61]	✓			✓	Standard drinks	Negative effect	Photorealistic drink images
Dry Days (2020; United Kingdom) [55]	✓	✓		✓	Standard drinks	Positive effect	—
Drynk (2020; Ireland) [49]	✓			✓	Standard drinks	—	BAC calculator
Simple Alcohol Unit Tracker (2020; United Kingdom) [57]	✓				Standard drinks	Negative effect	—
Try Dry (2020; United Kingdom) [52]	✓	✓		✓	Standard drinks	Physiology or nutrition; positive effect	AUDIT-C ^l
Mobile app: SD counter							
Wise Drinking (2019; France) [48]	✓	✓		✓	Standard drinks	Physiology or nutrition	Available in Russian

Tool name (year of last update ^b ; country)	Animation features			Adjust drinks ^c (n=13)	User feedback		Extra features
	<i>Drinks</i> ^d (n=21)	<i>Vessels</i> ^e (n=9)	<i>Pour</i> ^f (n=2)		Unit of consumption ^g	Additional feedback ^h	
Web-based tool: 1-time assessment of risky drinking							
DrinkCoach Alcohol Test (not provided; United Kingdom) [56]	✓	✓		Standard drinks; AUDIT risk score	Physiology or nutrition	Visually enhanced AUDIT; linked to the DrinkCoach mobile tool	
HSE Self-assessment tool (2019; Ireland) [64]	✓			Standard drinks; AUDIT risk score	Physiology or nutrition	Visually enhanced AUDIT	
Web-based tool: SD counter							
Drinkaware Drinks Calculator (2020; Ireland) [65]	✓			Standard drinks	Physiology or nutrition; negative effect	Drink selection depends on chosen drinking context	

^aDAT: digital assessment tool.

^bAt time of data extraction (2020).

^cNonvisually adjust drink characteristics.

^dChoose drinks from visual selection.

^eChoose vessels from visual selection.

^fSimulated interactive *pouring* of drinks.

^gStandard drinks, alcohol quantity (pure ethanol consumed [eg, in g or L]), and AUDIT risk score.

^hPhysiology- or nutrition-related feedback (eg, calories, ingested sugar, alcohol quantity equivalent in volume of beer or vodka, *burger equivalent*, exercise time to burn calories, typical symptoms at intoxication level, time until sober); negative effect of consumption (eg, money spent, heavy drinking days, drinking days per week); positive effect of reduced consumption (eg, money saved, sober days).

ⁱAUDIT: Alcohol Use Disorders Identification Test.

^jNot available.

^kBAC: blood alcohol concentration.

^lAUDIT-C: Alcohol Use Disorders Identification Test–Consumption.

Use of Animation

The 3 animation features defined in the eligibility criteria represent different levels of animation complexity. The distribution of these animation features in the identified alcohol DATs is summarized in [Textbox 1](#).

With the exception of 9% (2/22) of tools using photorealistic images [47,61], all tools presented a selection of abstract and often simplified drink icons. Examples of the assessment screens are shown in [Figure 2](#). More than half of the tools (12/22, 55%) offered only 1 animation feature (*selection of a drink or a beverage type from a number of visually displayed options*). The number of drinks to choose from differed considerably. Tools with fewer choices (<10 drink icons) [46,49,50,57,63,64,66] did not allow for any individualization of the chosen drink, whereas tools with more choices (16-29 drink icons) [47,56,61] enabled the user to individually adjust certain drink characteristics, including standard units of alcohol [56], drinking vessel size [61], volume consumed [47], and alcohol content of the consumed beverage. The *Know Your Units* tool [53] featured a *virtual bar* animation [53], allowing the user to drag a predefined drink icon from a shelf onto a bar table, where it was emptied out, accompanied by a beverage-specific sound animation.

Most other tools (8/22, 36%) relied on a 2-step process to describe consumed drinks. After choosing a beverage category,

users could choose their glass or bottle from a beverage-specific selection. In the group of mobile tools, the choice of available beverage categories and vessels per category differed from basic (3-7 beverage categories; ≤3 vessel icons per category) [48,59,62] to detailed (6-7 beverage categories; 4-10 vessel icons per category) [52-55]. All but 25% (2/8) of these tools [48,59] displayed all individualization steps on 1 overview screen (eg, [Figure 2, Try Dry](#)). All tools allowed for nonvisual adjustment of beverage alcohol content [48,52-55,59] or consumed beverage quantity [59,62]. The *DrinkCoach* web-based tool [56] lets users choose from 12 beverage categories, as well as 3 to 7 vessels per beverage category, but did not allow for further adjustment of drink characteristics.

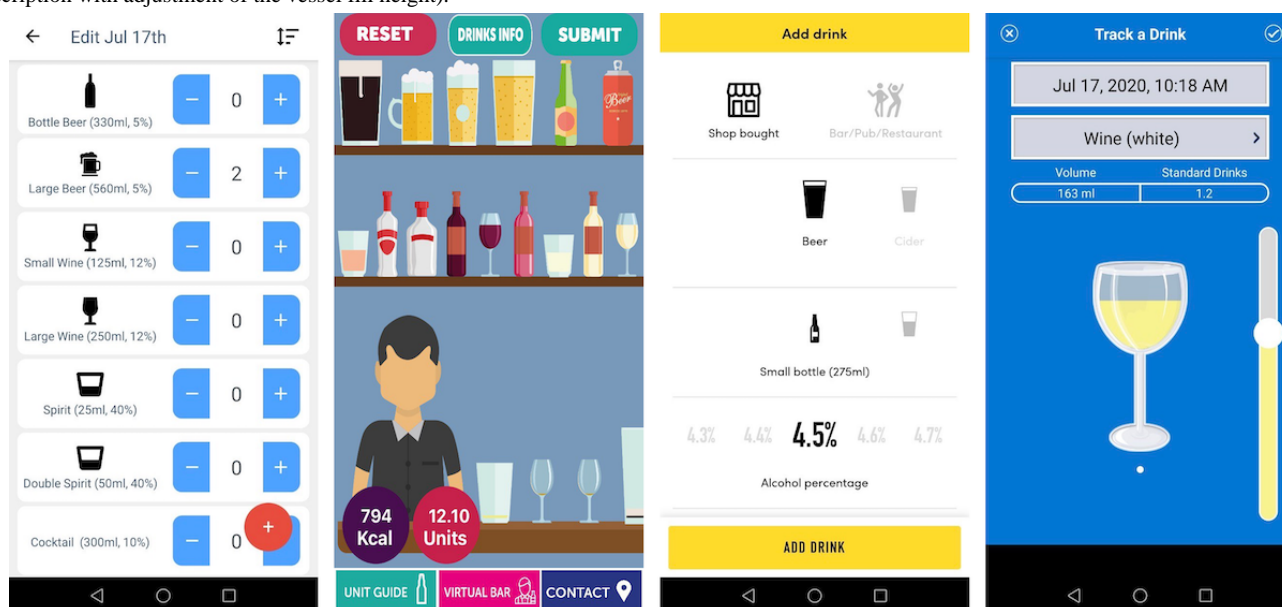
Only 9% (2/22) of the included tools featured the simulated interactive *pouring* of a drink; that is, continuously adjusting the beverage level displayed in a drinking vessel [58,60]. One of these tools used a nonchangeable standard vessel icon and a standard-colored beverage for the animation [60]. The other tool allowed users to choose the beverage and the vessel before pouring their drink ([Figure 2, Saying When](#)) [58]. The poured volume was displayed in real time during the pouring action, in milliliters as well as in standard drinks. The color of the liquid matched that of the chosen beverage. Further features to enhance the 3D character of the pouring experience, such as shadows, sound animations, or pouring-induced movement of the liquid surface were not identified.

Textbox 1. Identified animation features in mobile and web-based alcohol digital assessment tools.

Identified animation features

- Most of the included tools (21/22, 95%) offered the selection of a drink or a beverage type from a number of visually displayed options (1-step visual description).
- Less than half of the tools (9/22, 41%) additionally offered the selection of a drinking vessel from a number of visually displayed glasses and, in some instances, bottles (2-step visual description).
- Only 9% (2/22) of tools allowed the simulated interactive pouring of a drink; that is, continuously adjusting the beverage level displayed in the chosen drinking vessel.

Figure 2. Screenshots of drink input sections in mobile digital assessment tools quantifying alcohol consumption. From left to right: Simple Alcohol Unit Tracker [58] and Know Your Units [54] (both 1-step visual description); Try Dry [53] (2-step visual description); Saying When [59] (2-step visual description with adjustment of the vessel fill height).



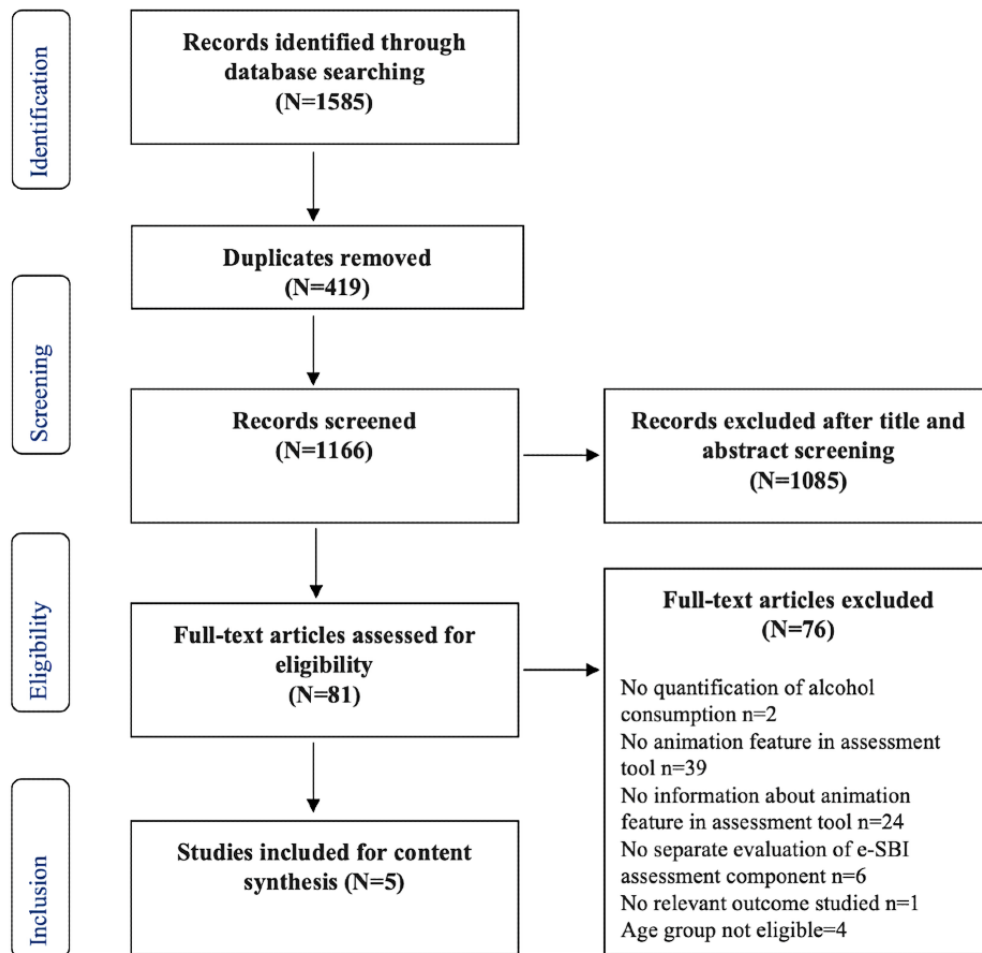
Part 2: Identification of Feasibility, Acceptability, and Validity Studies

Overview

A total of 1585 records were identified through a systematic literature review search in Embase, MEDLINE, PsycINFO, and

Web of Science (Figure 3). Removal of duplicates left 73.56% (1166/1585) of records for the title and abstract screening. In the random sample of 100 records screened by 2 reviewers, the agreement was 92% for inclusion decisions. Of the 81 full-text articles assessed for eligibility, 5 (6%) met the inclusion criteria.

Figure 3. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of study selection. eSBI: electronic screening and brief intervention.



Study Characteristics

All included studies used a 1-arm study design with convenience sampling to explore the feasibility, acceptability, or validity of the alcohol DAT in question, or several of these concepts (Table 2). Of the 5 studies, 3 (60%) were conducted in Australia [69-71], 1 (20%) in Canada [72], and 1 (20%) study used a combined sample of participants recruited in Canada and Switzerland [73]. Data collection took place between 2015 and 2017; 40% (2/5) of papers [70,71] did not report the period of data collection.

Participants were recruited at primary health care and addiction centers [69,71] and through researcher networks [70], advertisements on university campus [70,72], social media, and internet forums [70,73]. Of 5 studies, 2 (40%) recruited participants from the general population [70,72]; 20% (n=1) of studies focused on adults with risky alcohol use [73]; and 40% (n=2) of studies used quotas to include nondrinkers, nondependent drinkers, and dependent drinkers [69,71]. Sample sizes ranged from 50 [72] to 671 [70] participants, with balanced proportions of men and women.

Table 2. Overview of included studies (N=5).

Reference	Country; time of data collection (tested alcohol DAT ^a)	Study sample (age in years)	Recruitment	Main findings on acceptability and criterion or convergent validity
Acceptability and feasibility studies				
Lee et al [69]	Australia; 2016-2017 (Grog Survey app)	246 patients (18-78) with and without problematic alcohol use; 5 field research assistants (— ^b)	Primary health care and addiction center	<ul style="list-style-type: none"> Acceptability: 97% of patients rated alcohol DAT as easy to use or okay to use (rather than hard to use); staff suggested a high potential for the app to be used in primary health care settings, noted that participants appeared engaged and required minimal assistance
Bertholet et al [73]	Switzerland and Canada; 2015 (Alcoquizz)	130 participants (mean 32.8, SD 10) with problematic alcohol use	Social media and internet forums	<ul style="list-style-type: none"> Acceptability: Low self-reported frequency of alcohol DATs use during the 3-month study period (only 53.6% of participants reported using it more than once); moderate rating for appreciation and usefulness of the alcohol DAT (mean 6/10 points, IQR 5-8)
Validation studies				
Lee et al [71]	Australia; 2019 ^c (Grog Survey app)	238 participants (18-78) with and without problematic alcohol use	Primary health care and addiction centers	<ul style="list-style-type: none"> Criterion and convergent validity: Moderate (Spearman correlation between alcohol DAT and clinical interview for consumption quantity: $r=0.68$; $P<.01$); compared with interviews, alcohol DAT recorded higher numbers of standard drinks consumed per drinking occasion (median 17.0, IQR 10.5-27.9 and median 15.4, IQR 9.6-23.2) Criterion validity: Equal or better correlation of the presence of self-reported withdrawal tremors with the self-reported quantity of alcohol consumption in the alcohol DAT ($r=0.40$; $P<.05$) than with consumption estimate in the clinical interview ($r=0.32$; $P<.05$)
Poulton et al [70]	Australia; 2018 ^c (CNLab-A)	671 participants (16-56) with unknown alcohol use	Researcher networks and social media and internet forums	<ul style="list-style-type: none"> Convergent validity: Acceptable or high, with a significantly higher percentage of drinking days ($P=.007$) and total alcohol intake ($P<.001$) assessed by EMA^d alcohol DAT compared with 21-day TLFB^e; alcohol DAT recorded significantly higher hourly alcohol intake compared with AUQ^f ($P=.002$); no significant difference between AUQ and DAT in estimated weekly average consumption ($P=.13$)
Vanderlee et al [72]	Canada; 2016 (Beverage Frequency Questionnaire)	50 participants (16-30) with unknown alcohol use	Advertisement on university campus	<ul style="list-style-type: none"> Convergent validity: High correlation with 7dFR^g for number of drinks (Pearson $r=0.58$; $P<.001$) and consumed volume ($r=0.78$; $P<.001$) Acceptability: Good comprehensiveness assessed through cognitive interviewing (78% of participants reported no trouble in selecting a beverage image).

^aDAT: digital assessment tool.

^bNot available.

^cYear of study, as the year of data collection is not available.

^dEMA: ecological momentary assessment.

^eTLFB: Alcohol Timeline Followback.

^fAUQ: Alcohol Use Questionnaire.

^g7dFR: 7-day food record.

Characteristics of the Evaluated Tools

The evaluated alcohol DATs included 3 mobile apps and 1 web-based tool (Table 3), none of which had been identified in

this review through the systematic app store and web search. Only 25% (1/4) of tools were publicly available in the German or Russian app stores [73]. They were designed to screen for

risky alcohol use and collect consumption data at the population level [69,71], enable real-time assessment of alcohol intake [70], deliver a program to reduce drinking [73], and conduct epidemiological research [72].

Of the 4 tools, 2 (50%) presented a low number of visually displayed drink choices (<10 drink icons) [70,73]; 25% (1/4)

of tools offered the additional choice of a drinking vessel (16 vessel icons in 4 alcoholic beverage categories) [72]. The *Grog Survey app* offered a wide range of region- and culture-specific beverages and drinking vessels and the additional feature of *pouring* a drink [71,75].

Table 3. Scientifically evaluated DATs^a quantifying alcohol consumption: overview of core characteristics.

Tool name (year of study)	Animation features			Adjust drinks ^b	User feedback		Extra features
	Drinks ^c	Vessels ^d	Pour ^e		Unit of consumption ^f	Additional feedback ^g	
eSBI^h for problematic alcohol use (mobile app)							
Alcooquizz (2017) [73]	✓				Risk score	Physiology or nutrition; negative effect	Comparison to reference group
Ecological momentary assessment alcohol DAT (mobile app)							
CNLab-A (2018) [70]	✓			✓	N/A ⁱ	— ^j	—
One-time assessment of risky drinking (mobile app)							
Grog Survey app (2019) [71,74]	✓	✓	✓		AUDIT ^k risk score	—	Visualizations partly use user-generated drinks
One-time alcohol consumption assessment for epidemiological research (web-based)							
Beverage Frequency Questionnaire (2018) [72]	✓	✓			N/A	—	Also assesses consumption of nonalcoholic drinks

^aDAT: digital assessment tool.

^bNonvisually adjust drink characteristics.

^cChoose drinks from visual selection.

^dChoose vessels from visual selection.

^eSimulated interactive *pouring* of drinks.

^fStandard drinks, alcohol quantity (pure ethanol consumed [eg, in g or L]), AUDIT risk score, and DAT designed for epidemiological research, did not report the results to the user.

^gPhysiology or nutrition-related feedback (eg, calories, ingested sugar, and alcohol quantity equivalent in volume of beer or vodka, *burger equivalent*, exercise time to burn calories, typical symptoms at intoxication level, and time until sober); negative effect of consumption (eg, money spent, heavy drinking days, and drinking days per week); positive effect of reduced consumption (eg, money saved; sober days)

^heSBI: electronic screening and brief intervention.

ⁱN/A: not applicable; DAT designed for epidemiological research, did not report the results to the user.

^jNot available.

^kAUDIT: Alcohol Use Disorders Identification Test.

Findings Regarding Acceptability and Feasibility

Of the identified 5 studies, 2 (40%) focused on the acceptability and feasibility of the evaluated alcohol DAT (Table 2) [69,73]. Both used participant rating scales, rating *appreciation* and *usefulness* [73] and *ease of use*, respectively [69]. One of the tools, which was offered to study participants to be used at their discretion during a 3-month period, recorded the self-reported frequency of use [73]. In the second study, conducted in a health care setting, quantitative and qualitative staff observations were taken into account.

User evaluations of alcohol DATs were moderate to favorable. The animation-using personal feedback module of the first tool received an average participant rating of 6/10 in both the *appreciation* and the *usefulness* scales [73]. However, the self-reported frequency of use was low. The second tool was

rated as *easy to use* or *okay to use* rather than *hard to use* by 97% of the study participants. Staff observations concluded that it could be completed with or without minimal assistance across different age groups [69].

Findings Regarding Validity

In total, 60% (3/5) of studies aimed to explore the validity of the respective alcohol DAT [70-72]. Established assessment methods to quantify alcohol consumption, such as clinical interviews or the alcohol Timeline Followback questionnaire, were used as comparators. One of the studies additionally evaluated the correlation between physical signs of addiction and the self-reported quantity of alcohol consumption [71].

The reported convergent validity was moderate in one of the studies [71] and moderate to high in a second study [72]. In 40% (2/5) of studies, the alcohol DAT recorded higher alcohol

consumption than the established assessment method [70,72]. In one case, comparing an alcohol DAT designed for real-time drinking assessment with a 21-day retrospective assessment, this difference was statistically significant for the percentage of drinking days and the total alcohol intake but not for the number of heavy drinking occasions [70]. In the other study, the number of standard drinks consumed per drinking occasion did not significantly differ between the alcohol DAT and the established assessment method [71]. Furthermore, consumption estimates recorded in the alcohol DAT predicted physical signs of addiction as good or better than a clinical interview [71].

Discussion

Principal Findings

This systematic review is the first on DATs using animation features to support the quantitative assessment of alcohol consumption, a novel approach in the emerging field of digital health. Only 9% (2/22) of the alcohol DATs identified in part 1 of the review used animation in the sense of dynamically animated images that can be modified through user interaction (*pouring* a drink). Most animation features were implemented in a simplistic manner and did not exploit the full visualization potential of the available technology. The addition of dynamic visual hints, such as foam, bubbles, or visible movement of the beverage, could potentially help users recall their drinking habits in greater detail, which is thought to enhance the accuracy of reporting [76,77]. The results indicate that these features remain underused and that there is ample room for exploration and development.

In the identified alcohol DATs, relevant information regarding the responsible organization, scientific background, and development process was often incomplete or unavailable, which prevented a well-founded quality assessment. A larger degree of transparency is urgently required to fully exploit the potential of animation-using alcohol DATs. Similarly, none of the included tools provided information on the cultural or regional adaptation of the offered beverages and drinking vessels [15,78] or the approach and data sources used for this process. Additional searches with Russian search terms, aiming to identify relevant Russian-language alcohol DAT and gauge the potential of repeating the systematic search in additional languages in the future, showed a high availability of English-language apps in Russian app stores. However, they did not yield evidence of efforts to account for different cultural contexts in different language versions of the same app. Moreover, to the best of our knowledge, none of the tools identified in the first part of the review had been scientifically evaluated, underlining the lack of evidence for animation-using alcohol DATs publicly available in app stores [79].

The second part of this review identified 5 exploratory studies on the feasibility, acceptability, and validity of 4 animation-using alcohol DATs. These data showed fair to high convergent validity between established consumption assessment methods and animation-using alcohol DATs, whereas some alcohol DATs were shown to record higher quantities of alcohol consumption than the established measure. Considering the worldwide underestimation of self-reported alcohol consumption

[22-24], these results could arguably be interpreted as a sign of improved assessment accuracy [21]. Animation-using alcohol DATs might thus contribute to reducing the well-known bias of standard surveys.

Strengths and Limitations

To not miss any relevant alcohol DAT using interactive visualizations, as opposed to purely text-based quantitation tools, a broad definition of the term *animation* was chosen, encompassing any apps that allow an image-based interaction with the user to quantify personal alcohol consumption. In part 1, systematic searches were conducted in Android and iOS app stores and via the Google search engine. These sources do not provide access to tools that are published in smaller stores, such as Amazon App Store, Samsung Apps, or Windows Store, or on open-source platforms, such as Github [80]. Alcohol DATs developed for health institutions or researchers may also have been missed, as they often use ways of dissemination not covered in this review [81]. We restricted our search to apps available free of charge. More than 95% of the apps in the Android market [41] and >99% of all downloaded and installed apps [42] are estimated to meet this criterion.

The specific limitations of part 1 stem from the characteristics of nonscientific search engines and app stores as search environments. The providers of the platforms searched for this review do not disclose their search algorithms [38,82]. Search parameters, such as language and region settings and customization based on previous search behavior, are known to influence the choice and order of results, reducing the replicability of searches. Copyright regulations and the differing contents of national app stores further limit the selection of apps available for review. Separate searches in all available national app stores would not have been feasible with the available resources. On the basis of the team's locations, we searched the German and Russian app stores, which showed a high content overlap with US and UK app stores [39]. The searches allowed for the identification of alcohol DATs from several countries. Obtaining true global or regional representativeness is beyond the scope of this review.

Moreover, digital app stores can be considered very unstable sources of information. Their contents change quickly over time and although for research articles stable identifiers, such as the digital object identifier number, have been developed, so far, there is no equivalent for mobile apps. Analyzing app store data through systematic searches is a relatively novel approach; thus, accepted reporting guidelines are not yet available. However, first recommendations have been developed [38,82], which guided the reporting in this review.

Many studies identified in part 2 focused on the evaluation of electronic SBI programs to reduce alcohol consumption and provided no information on the use of animation features in the alcohol DAT that was tested. If this information could be obtained through a web search, the study was excluded from the review. Therefore, it is possible that some studies evaluating animation-using alcohol DATs were falsely excluded.

Comparison With Prior Work and Future Research Perspectives

Today, >318,000 health apps are available on the app stores [83], most of which are not recognizably evidence based [79,84]. Many apps, especially in the field of alcohol use, even promote harmful behavior [85]. Efforts have been undertaken to develop frameworks for app quality evaluation [86,87], as well as provide systematic evaluations of the apps available in specified health fields [88-90]. There are promising data on both the efficacy of health apps to reduce harmful alcohol consumption [32,33] and on the effect of interactive elements and gamification in health apps [91-93]. This review adds a separate evaluation of the availability and effects of interactive animation features on alcohol DAT.

Further research is needed to evaluate the differences between regional app markets within and beyond the English- and Russian-language markets. To facilitate the implementation of animation-using alcohol DAT in existing health care systems, target group-specific evaluations, analyzing the perspective of different age groups, and professional versus patient experiences with animation-using alcohol DAT would be highly valuable. The cognitive and psychological mechanisms underlying the effects of animation features also warrant further evaluation. For future research and tool development, the field of alcohol assessment might benefit from deepening the dialog with

nutritional and dietary studies that have already started the development of interactive tools using more elaborate animation features that showed high validity and user-friendliness [30,94,95].

Conclusions

Research in the field of DATs is rapidly advancing. This is especially true for the area of mental health assessment tools, platforms, and resources and seems particularly urgent in light of the current COVID-19 pandemic [96-98]. By facilitating the collection of internationally comparable data as part of population-based surveys and improving the delivery of electronic SBIs for hazardous and harmful alcohol use, animation-using alcohol DATs might contribute to reducing alcohol-attributable health burden in the future. However, the potential of using animation features for the quantification of individual alcohol intake in DATs has not been fully exploited to date and has received little scientific attention. Further research is needed to explore the extent to which such features could improve the accuracy and user-friendliness of the assessment and identify the underlying mechanisms. However, although mostly using nondynamic animation features and often deficient in scientific background information, first animation-using alcohol DATs are available in app stores and on the web, and the explorative study data generated so far support their novel approach.

Acknowledgments

Financial support for the project was provided through a grant from the Government of Germany and the Russian Federation in the context of the World Health Organization (WHO) European Office for the Prevention and Control of Noncommunicable Diseases. These funding sources had no role in the design of this study, analyses, interpretation of the data, or decision to submit the results. Carina Ferreira-Borges and Maria Neufeld are staff members of the WHO Regional Office for Europe. The authors alone are responsible for the views expressed in this publication, and these do not necessarily represent the decisions or stated policy of the WHO.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Keyword set (Ovid search).

[DOC File, 31 KB - [jmir_v24i3e28927_app1.doc](#)]

Multimedia Appendix 2

Included digital assessment tools.

[PDF File (Adobe PDF File), 350 KB - [jmir_v24i3e28927_app2.pdf](#)]

References

1. GBD 2016 Alcohol and Drug Use Collaborators. The global burden of disease attributable to alcohol and drug use in 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Psychiatry* 2018 Dec;5(12):987-1012 [FREE Full text] [doi: [10.1016/S2215-0366\(18\)30337-7](#)] [Medline: [30392731](#)]
2. Heather N. Can screening and brief intervention lead to population-level reductions in alcohol-related harm? *Addict Sci Clin Pract* 2012;7:15 [FREE Full text] [doi: [10.1186/1940-0640-7-15](#)] [Medline: [23186309](#)]
3. Angus C, Thomas C, Anderson P, Meier PS, Brennan A. Estimating the cost-effectiveness of brief interventions for heavy drinking in primary health care across Europe. *Eur J Public Health* 2017 Apr 01;27(2):345-351. [doi: [10.1093/eurpub/ckw122](#)] [Medline: [27558943](#)]

4. Tansil KA, Esser MB, Sandhu P, Reynolds JA, Elder RW, Williamson RS, Community Preventive Services Task Force. Alcohol electronic screening and brief intervention: a community guide systematic review. *Am J Prev Med* 2016 Nov;51(5):801-811 [FREE Full text] [doi: [10.1016/j.amepre.2016.04.013](https://doi.org/10.1016/j.amepre.2016.04.013)] [Medline: [27745678](https://pubmed.ncbi.nlm.nih.gov/27745678/)]
5. SAFER: preventing and reducing alcohol-related harms. World Health Organization. 2018. URL: https://www.who.int/substance_abuse/safer/msb_safer_framework.pdf?ua=1 [accessed 2021-12-13]
6. US Preventive Services Task Force, Curry SJ, Krist AH, Owens DK, Barry MJ, Caughey AB, et al. Screening and behavioral counseling interventions to reduce unhealthy alcohol use in adolescents and adults: US preventive services task force recommendation statement. *J Am Med Assoc* 2018 Nov 13;320(18):1899-1909. [doi: [10.1001/jama.2018.16789](https://doi.org/10.1001/jama.2018.16789)] [Medline: [30422199](https://pubmed.ncbi.nlm.nih.gov/30422199/)]
7. Anderson P, Wojnar M, Jakubczyk A, Gual A, Reynolds J, Segura L, et al. Managing alcohol problems in general practice in Europe: results from the European ODHIN survey of general practitioners. *Alcohol Alcoholism* 2014;49(5):531-539 [FREE Full text] [doi: [10.1093/alcalc/agu043](https://doi.org/10.1093/alcalc/agu043)] [Medline: [25031247](https://pubmed.ncbi.nlm.nih.gov/25031247/)]
8. Anderson P, Braddick F, Reynolds J, Gual A. Alcohol policy in Europe: evidence from AMPHORA. 2nd ed. In: *Alcohol Measures for Public Health Research Alliance (AMPHORA)*. Brussels, Belgium: European Commission; 2012.
9. Rosário F, Santos MI, Angus K, Pas L, Ribeiro C, Fitzgerald N. Factors influencing the implementation of screening and brief interventions for alcohol use in primary care practices: a systematic review using the COM-B system and Theoretical Domains Framework. *Implement Sci* 2021 Jan 07;16(1):6 [FREE Full text] [doi: [10.1186/s13012-020-01073-0](https://doi.org/10.1186/s13012-020-01073-0)] [Medline: [33413487](https://pubmed.ncbi.nlm.nih.gov/33413487/)]
10. Sierosławski J, Foster J, Moskalewicz J. Survey of European drinking surveys. Alcohol survey experiences of 22 European countries. *Drugs Edu Prevent Policy* 2013 May 31;20(5):383-398. [doi: [10.3109/09687637.2013.797381](https://doi.org/10.3109/09687637.2013.797381)]
11. Greenfield TK. Ways of measuring drinking patterns and the difference they make: experience with graduated frequencies. *J Subst Abuse* 2000;12(1-2):33-49. [doi: [10.1016/s0899-3289\(00\)00039-0](https://doi.org/10.1016/s0899-3289(00)00039-0)] [Medline: [11288473](https://pubmed.ncbi.nlm.nih.gov/11288473/)]
12. Dawson DA. Methodological issues in measuring alcohol use. *Alcohol Res Health* 2003;27(1):18-29 [FREE Full text] [Medline: [15301397](https://pubmed.ncbi.nlm.nih.gov/15301397/)]
13. Babor T, Higgins-Biddle J, Saunders J, Monteiro M. The alcohol use disorders identification test: guidelines for use in primary care, second edition. World Health Organization. 2001. URL: <https://apps.who.int/iris/handle/10665/67205> [accessed 2021-12-13]
14. Bush K, Kivlahan DR, McDonnell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Arch Intern Med* 1998 Sep 14;158(16):1789-1795. [doi: [10.1001/archinte.158.16.1789](https://doi.org/10.1001/archinte.158.16.1789)] [Medline: [9738608](https://pubmed.ncbi.nlm.nih.gov/9738608/)]
15. Kalinowski A, Humphreys K. Governmental standard drink definitions and low-risk alcohol consumption guidelines in 37 countries. *Addiction* 2016 Jul 13;111(7):1293-1298. [doi: [10.1111/add.13341](https://doi.org/10.1111/add.13341)] [Medline: [27073140](https://pubmed.ncbi.nlm.nih.gov/27073140/)]
16. Babor T, Robaina K. The Alcohol Use Disorders Identification Test (AUDIT): a review of graded severity algorithms and national adaptations. *Int J Alcohol Drug Res* 2016 Jul 19;5(2):17-24. [doi: [10.7895/ijadr.v5i2.222](https://doi.org/10.7895/ijadr.v5i2.222)]
17. Kerr W, Stockwell T. Understanding standard drinks and drinking guidelines. *Drug Alcohol Rev* 2012 Mar;31(2):200-205 [FREE Full text] [doi: [10.1111/j.1465-3362.2011.00374.x](https://doi.org/10.1111/j.1465-3362.2011.00374.x)] [Medline: [22050262](https://pubmed.ncbi.nlm.nih.gov/22050262/)]
18. Sprague DJ, Vinson DC. Patient perceptions of risky drinking: knowledge of daily and weekly low-risk guidelines and standard drink sizes. *Subst Abuse* 2017;38(3):253-256. [doi: [10.1080/08897077.2015.1048922](https://doi.org/10.1080/08897077.2015.1048922)] [Medline: [26155748](https://pubmed.ncbi.nlm.nih.gov/26155748/)]
19. Boniface S, Kneale J, Shelton N. Actual and perceived units of alcohol in a self-defined "usual glass" of alcoholic drinks in England. *Alcohol Clin Exp Res* 2013 Jun 20;37(6):978-983 [FREE Full text] [doi: [10.1111/acer.12046](https://doi.org/10.1111/acer.12046)] [Medline: [23278164](https://pubmed.ncbi.nlm.nih.gov/23278164/)]
20. Welsh C, Earley K, Delahanty J, Wright K, Berens T, Williams A, et al. Residents' knowledge of standard drink equivalents: implications for screening and brief intervention for at-risk alcohol use. *Am J Addict* 2014;23(2):194-196. [doi: [10.1111/j.1521-0391.2013.12080.x](https://doi.org/10.1111/j.1521-0391.2013.12080.x)] [Medline: [24112850](https://pubmed.ncbi.nlm.nih.gov/24112850/)]
21. Del Boca FK, Darkes J. The validity of self-reports of alcohol consumption: state of the science and challenges for research. *Addiction* 2003 Dec;98 Suppl 2:1-12. [doi: [10.1046/j.1359-6357.2003.00586.x](https://doi.org/10.1046/j.1359-6357.2003.00586.x)] [Medline: [14984237](https://pubmed.ncbi.nlm.nih.gov/14984237/)]
22. Boniface S, Shelton N. How is alcohol consumption affected if we account for under-reporting? A hypothetical scenario. *Eur J Public Health* 2013 Dec;23(6):1076-1081. [doi: [10.1093/eurpub/ckt016](https://doi.org/10.1093/eurpub/ckt016)] [Medline: [23444427](https://pubmed.ncbi.nlm.nih.gov/23444427/)]
23. Probst C, Shuper PA, Rehm J. Coverage of alcohol consumption by national surveys in South Africa. *Addiction* 2017 Apr 20;112(4):705-710. [doi: [10.1111/add.13692](https://doi.org/10.1111/add.13692)] [Medline: [27859902](https://pubmed.ncbi.nlm.nih.gov/27859902/)]
24. Kilian C, Manthey J, Probst C, Brunborg GS, Bye EK, Ekholm O, et al. Why is per capita consumption underestimated in alcohol surveys? Results from 39 surveys in 23 European countries. *Alcohol Alcohol* 2020 Aug 14;55(5):554-563 [FREE Full text] [doi: [10.1093/alcalc/agua048](https://doi.org/10.1093/alcalc/agua048)] [Medline: [32491170](https://pubmed.ncbi.nlm.nih.gov/32491170/)]
25. Campbell N, Ali F, Finlay AY, Salek SS. Equivalence of electronic and paper-based patient-reported outcome measures. *Qual Life Res* 2015 Aug;24(8):1949-1961. [doi: [10.1007/s11136-015-0937-3](https://doi.org/10.1007/s11136-015-0937-3)] [Medline: [25702266](https://pubmed.ncbi.nlm.nih.gov/25702266/)]
26. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. *Value Health* 2008;11(2):322-333. [doi: [10.1111/j.1524-4733.2007.00231.x](https://doi.org/10.1111/j.1524-4733.2007.00231.x)] [Medline: [18380645](https://pubmed.ncbi.nlm.nih.gov/18380645/)]

27. Harris SK, Knight JR, Van Hook S, Sherritt L, Brooks TL, Kulig JW, et al. Adolescent substance use screening in primary care: validity of computer self-administered versus clinician-administered screening. *Subst Abus* 2016;37(1):197-203 [FREE Full text] [doi: [10.1080/08897077.2015.1014615](https://doi.org/10.1080/08897077.2015.1014615)] [Medline: [25774878](https://pubmed.ncbi.nlm.nih.gov/25774878/)]
28. Harris SK, Knight JR. Putting the screen in screening: technology-based alcohol screening and brief interventions in medical settings. *Alcohol Res* 2014;36(1):63-79 [FREE Full text] [Medline: [26259001](https://pubmed.ncbi.nlm.nih.gov/26259001/)]
29. Sawares AS, Shen N, Xue Y, Abi-Jaoude A, Wiljer D. The impact of mobile apps on alcohol use disorder: a systematic review protocol. *JMIR Res Protoc* 2017 Apr 04;6(4):e49 [FREE Full text] [doi: [10.2196/resprot.6975](https://doi.org/10.2196/resprot.6975)] [Medline: [28377366](https://pubmed.ncbi.nlm.nih.gov/28377366/)]
30. Illner A, Freisling H, Boeing H, Huybrechts I, Crispim SP, Slimani N. Review and evaluation of innovative technologies for measuring diet in nutritional epidemiology. *Int J Epidemiol* 2012 Aug;41(4):1187-1203 [FREE Full text] [doi: [10.1093/ije/dys105](https://doi.org/10.1093/ije/dys105)] [Medline: [22933652](https://pubmed.ncbi.nlm.nih.gov/22933652/)]
31. Amoutzopoulos B, Steer T, Roberts C, Cade JE, Boushey CJ, Collins CE, et al. Traditional methods new technologies - dilemmas for dietary assessment in large-scale nutrition surveys and studies: a report following an international panel discussion at the 9th International Conference on Diet and Activity Methods (ICDAM9), Brisbane, 3 September 2015. *J Nutr Sci* 2018;7:e11 [FREE Full text] [doi: [10.1017/jns.2018.4](https://doi.org/10.1017/jns.2018.4)] [Medline: [29686860](https://pubmed.ncbi.nlm.nih.gov/29686860/)]
32. Meredith SE, Alessi SM, Petry NM. Smartphone applications to reduce alcohol consumption and help patients with alcohol use disorder: a state-of-the-art review. *Adv Health Care Technol* 2015;1:47-54 [FREE Full text] [doi: [10.2147/AHCT.S65791](https://doi.org/10.2147/AHCT.S65791)] [Medline: [27478863](https://pubmed.ncbi.nlm.nih.gov/27478863/)]
33. Wieczorek L, Klingemann JI. Mobile applications used to limit alcohol consumption – a literature review. *Alcohol Drug Addict* 2020;33(1):43-64. [doi: [10.5114/ain.2020.95979](https://doi.org/10.5114/ain.2020.95979)]
34. Attwood S, Parke H, Larsen J, Morton KL. Using a mobile health application to reduce alcohol consumption: a mixed-methods evaluation of the drinkaware track and calculate units application. *BMC Public Health* 2017 Dec 17;17(1):394 [FREE Full text] [doi: [10.1186/s12889-017-4358-9](https://doi.org/10.1186/s12889-017-4358-9)] [Medline: [28511698](https://pubmed.ncbi.nlm.nih.gov/28511698/)]
35. Garnett C, Oldham M, Angus C, Beard E, Burton R, Field M, et al. Evaluating the effectiveness of the smartphone app, Drink Less, compared with the NHS alcohol advice webpage, for the reduction of alcohol consumption among hazardous and harmful adult drinkers in the UK at 6-month follow-up: protocol for a randomised controlled trial. *Addiction* 2021 Feb;116(2):412-425 [FREE Full text] [doi: [10.1111/add.15287](https://doi.org/10.1111/add.15287)] [Medline: [33067856](https://pubmed.ncbi.nlm.nih.gov/33067856/)]
36. PROSPERO: International prospective register of systematic reviews. National Institute for Health Research. URL: <https://www.crd.york.ac.uk/prospere/> [accessed 2021-12-13]
37. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br Med J* 2021 Mar 29;372:n71 [FREE Full text] [doi: [10.1136/bmj.n71](https://doi.org/10.1136/bmj.n71)] [Medline: [33782057](https://pubmed.ncbi.nlm.nih.gov/33782057/)]
38. Grainger R, Devan H, Sangelaji B, Hay-Smith J. Issues in reporting of systematic review methods in health app-focused reviews: a scoping review. *Health Informatics J* 2020 Dec 11;26(4):2930-2945 [FREE Full text] [doi: [10.1177/1460458220952917](https://doi.org/10.1177/1460458220952917)] [Medline: [32914696](https://pubmed.ncbi.nlm.nih.gov/32914696/)]
39. *Winners & Losers in the Global App Economy*. Farnham, Surrey, United Kingdom: Caribou Digital Publishing; 2016.
40. Huang Z, Soljak M, Boehm BO, Car J. Clinical relevance of smartphone apps for diabetes management: a global overview. *Diabetes Metab Res Rev* 2018 May;34(4):e2990. [doi: [10.1002/dmrr.2990](https://doi.org/10.1002/dmrr.2990)] [Medline: [29431916](https://pubmed.ncbi.nlm.nih.gov/29431916/)]
41. Number of available Android applications. AppBrain. URL: <https://www.appbrain.com/stats/free-and-paid-android-applications> [accessed 2021-12-13]
42. Wang H, Li H, Guo Y. Understanding the evolution of mobile app ecosystems: a longitudinal measurement study of Google play. In: *Proceedings of the The World Wide Web Conference.: Association for Computing Machinery, Inc; 2019 Presented at: WWW '19: The Web Conference; May 13 - 17, 2019; San Francisco CA USA p. 1988-1999.* [doi: [10.1145/3308558.3313611](https://doi.org/10.1145/3308558.3313611)]
43. Neufeld M, Ferreira-Borges C, Gil A, Manthey J, Rehm J. Alcohol policy has saved lives in the Russian Federation. *Int J Drug Policy* 2020 Jun;80:102636. [doi: [10.1016/j.drugpo.2019.102636](https://doi.org/10.1016/j.drugpo.2019.102636)] [Medline: [32417670](https://pubmed.ncbi.nlm.nih.gov/32417670/)]
44. WHO ASSIST Working Group. The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. *Addiction* 2002 Sep;97(9):1183-1194. [doi: [10.1046/j.1360-0443.2002.00185.x](https://doi.org/10.1046/j.1360-0443.2002.00185.x)] [Medline: [12199834](https://pubmed.ncbi.nlm.nih.gov/12199834/)]
45. Sobell LC, Sobell MB. Timeline follow-back. In: Litten RZ, Allen JP, editors. *Measuring Alcohol Consumption*. Totowa, NJ: Humana Press; 1992:41-72.
46. *ReduceYourDrinking - Reduce alcohol consumption successfully* [Mobile App]. CorporateCom GmbH. 2015. URL: <https://apps.apple.com/us/app/id943492616> [accessed 2021-07-08]
47. *AlcoExpert: alcohol tracker* [Mobile app]. Grey Media. 2019. URL: <https://play.google.com/store/apps/details?id=com.alcoexpert> [accessed 2021-07-08]
48. *Wise Drinking: let's be smart* [Mobile app]. Pernod Ricard. 2019. URL: <https://play.google.com/store/apps/details?id=com.pernodricard.wisedrinking&hl=gs&gl=US> [accessed 2021-07-08]
49. *Drynk alcohol tracker and insights* [Mobile app]. Syzible. 2020. URL: <https://play.google.com/store/apps/details?id=com.syzible.drynk&hl=de&gl=US> [accessed 2021-07-08]

50. alcCalc [Mobile App]. Noth Co. Ltd. 2014. URL: <https://apps.apple.com/us/app/alccalc-estimates-displays-alcohol-decomposition-time/id887557458> [accessed 2021-07-08]
51. Know Your Numbers (WessexAHSN) [Mobile app]. Network WAHS. 2017. URL: <https://play.google.com/store/apps/details?id=com.wessexahsn.knowyournumbers&hl=de&gl=US> [accessed 2021-07-08]
52. TRY DRY: the app for dry January and beyond [Mobile app]. Alcohol Change, UK. 2020. URL: <https://play.google.com/store/apps/details?id=uk.org.alcoholconcern.dryjanuar> [accessed 2021-07-08]
53. Know your units [Mobile app]. Alcohol And You, Northern Ireland. 2017. URL: <https://apps.apple.com/bn/app/know-your-units/id1260567821> [accessed 2021-07-08]
54. Drinks meter [Mobile app]. Global Drug Survey. 2020. URL: <https://play.google.com/store/apps/details?id=com.drinksmeter.alcohol> [accessed 2021-07-08]
55. Dry days by AlcoChange [Mobile app]. AlcoChange. 2020. URL: <https://apps.apple.com/us/app/dry-days-by-alcochange/id1062948603> [accessed 2021-07-08]
56. DrinkCoach [Mobile app]. Humankind. 2019. URL: https://play.google.com/store/apps/details?id=com.td.drinkcoach&hl=en_US&gl=US [accessed 2021-07-08]
57. Simple alcohol unit tracker [Mobile app]. DeveloperJam. App Store; 2020. URL: <https://play.google.com/store/apps/details?id=com.developerjam.drinklogger> [accessed 2021-07-08]
58. Saying when [Mobile App]. CAMH (Centre for Addiction and Mental Health). 2016. URL: <https://play.google.com/store/apps/details?id=com.sayingwhencamh> [accessed 2021-07-08]
59. MeSelfControl [Mobile App]. AppPeople. 2016. URL: <https://apps.apple.com/ca/app/meselfcontrol/id1051751592> [accessed 2021-07-08]
60. Alcohol meter [Mobile App]. Alcaraz J (ZaraclaJ). 2019. URL: https://play.google.com/store/apps/details?id=zaraclaj.alcohol_counter [accessed 2021-07-08]
61. DrinkControl: alcohol tracker [Mobile app]. E-protect. 2020. URL: <https://apps.apple.com/us/app/drinkcontrol-alcohol-tracker/id456207840> [accessed 2021-07-08]
62. Alcofy - Alcohol tracker and BAC drink calculator [Mobile app]. JMJ Dev. 2020. URL: <https://play.google.com/store/apps/details?id=com.jmstudio.alctrack> [accessed 2021-07-08]
63. Drink watch unit checker [Mobile app]. FASA - Henry McCrory. 2016. URL: <https://apps.apple.com/us/app/drink-watch-unit-checker/id1186722413> [accessed 2021-08-07]
64. HSE Self assessment tool [Web-based tool]. Health Service Executive, Ireland. URL: <https://www2.hse.ie/wellbeing/alcohol/self-assessment-tool/#Start> [accessed 2021-12-21]
65. Drinkaware - Drinks calculator [Web-based tool]. Alcohol Awareness Foundation Ireland T/A. URL: <https://www.drinkaware.ie/tools-resources/drinks-calculator> [accessed 2021-12-21]
66. Alcohol diary [Mobile App]. James Porter. 2019. URL: <https://apps.apple.com/us/app/alcohol-diary/id1437094610> [accessed 2021-07-08]
67. Luczak SE, Hawkins AL, Dai Z, Wichmann R, Wang C, Rosen IG. Obtaining continuous BrAC/BAC estimates in the field: a hybrid system integrating transdermal alcohol biosensor, Intellidrink smartphone app, and BrAC Estimator software tools. *Addict Behav* 2018 Aug;83:48-55 [FREE Full text] [doi: [10.1016/j.addbeh.2017.11.038](https://doi.org/10.1016/j.addbeh.2017.11.038)] [Medline: [29233567](https://pubmed.ncbi.nlm.nih.gov/29233567/)]
68. Alcohol Guidelines Review: summary of the proposed new guidelines. UK Chief Medical Officer. 2016. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/489795/summary.pdf [accessed 2021-12-21]
69. Lee KK, Conigrave JH, Al Ansari M, Wilson S, Perry J, Zheng C, et al. Acceptability and feasibility of a computer-based application to help Aboriginal and Torres Strait Islander Australians describe their alcohol consumption. *J Ethn Subst Abuse* 2021 Mar 19;20(1):16-33. [doi: [10.1080/15332640.2019.1579144](https://doi.org/10.1080/15332640.2019.1579144)] [Medline: [30887909](https://pubmed.ncbi.nlm.nih.gov/30887909/)]
70. Poulton A, Pan J, Bruns LR, Sinnott RO, Hester R. Assessment of alcohol intake: retrospective measures versus a smartphone application. *Addict Behav* 2018 Aug;83:35-41. [doi: [10.1016/j.addbeh.2017.11.003](https://doi.org/10.1016/j.addbeh.2017.11.003)] [Medline: [29128148](https://pubmed.ncbi.nlm.nih.gov/29128148/)]
71. Lee KSK, Conigrave JH, Callinan S, Wilson S, Room R, Perry J, et al. Asking about the last four drinking occasions on a tablet computer as a way to record alcohol consumption in Aboriginal and Torres Strait Islander Australians: a validation. *Addict Sci Clin Pract* 2019 May 01;14(1):15 [FREE Full text] [doi: [10.1186/s13722-019-0148-2](https://doi.org/10.1186/s13722-019-0148-2)] [Medline: [31039824](https://pubmed.ncbi.nlm.nih.gov/31039824/)]
72. Vanderlee L, Reid JL, White CM, Hobin EP, Acton RB, Jones AC, et al. Evaluation of the online Beverage Frequency Questionnaire (BFQ). *Nutr J* 2018 Aug 01;17(1):73 [FREE Full text] [doi: [10.1186/s12937-018-0380-8](https://doi.org/10.1186/s12937-018-0380-8)] [Medline: [30068342](https://pubmed.ncbi.nlm.nih.gov/30068342/)]
73. Bertholet N, Daeppen J, McNeely J, Kushnir V, Cunningham JA. Smartphone application for unhealthy alcohol use: a pilot study. *Subst Abuse* 2017;38(3):285-291. [doi: [10.1080/08897077.2017.1281860](https://doi.org/10.1080/08897077.2017.1281860)] [Medline: [28113039](https://pubmed.ncbi.nlm.nih.gov/28113039/)]
74. Lee K, Conigrave J, Wilson S, Perry J, Callinan S, Room R, et al. Short screening tools for risky drinking in Aboriginal and Torres Strait Islander Australians: modified AUDIT-C and a new approach. *Addict Sci Clin Pract* 2019 Jul 01;14(1):22 [FREE Full text] [doi: [10.1186/s13722-019-0152-6](https://doi.org/10.1186/s13722-019-0152-6)] [Medline: [31256762](https://pubmed.ncbi.nlm.nih.gov/31256762/)]
75. Lee KK, Wilson S, Perry J, Room R, Callinan S, Assan R, et al. Developing a tablet computer-based application ('App') to measure self-reported alcohol consumption in Indigenous Australians. *BMC Med Inform Decis Mak* 2018 Jan 15;18(1):8 [FREE Full text] [doi: [10.1186/s12911-018-0583-0](https://doi.org/10.1186/s12911-018-0583-0)] [Medline: [29334962](https://pubmed.ncbi.nlm.nih.gov/29334962/)]

76. McKenna H, Treanor C, O'Reilly D, Donnelly M. Evaluation of the psychometric properties of self-reported measures of alcohol consumption: a COSMIN systematic review. *Subst Abuse Treat Prev Policy* 2018 Feb 02;13(1):6 [FREE Full text] [doi: [10.1186/s13011-018-0143-8](https://doi.org/10.1186/s13011-018-0143-8)] [Medline: [29394950](https://pubmed.ncbi.nlm.nih.gov/29394950/)]
77. Greenfield TK, Kerr WC. Alcohol measurement methodology in epidemiology: recent advances and opportunities. *Addiction* 2008 Jul;103(7):1082-1099 [FREE Full text] [doi: [10.1111/j.1360-0443.2008.02197.x](https://doi.org/10.1111/j.1360-0443.2008.02197.x)] [Medline: [18422826](https://pubmed.ncbi.nlm.nih.gov/18422826/)]
78. Bräker AB, Soellner R. Alcohol drinking cultures of European adolescents. *Eur J Public Health* 2016 Aug;26(4):581-586. [doi: [10.1093/eurpub/ckw033](https://doi.org/10.1093/eurpub/ckw033)] [Medline: [27037330](https://pubmed.ncbi.nlm.nih.gov/27037330/)]
79. Weaver ER, Horyniak DR, Jenkinson R, Dietze P, Lim MS. "Let's get wasted!" and other apps: characteristics, acceptability, and use of alcohol-related smartphone applications. *JMIR Mhealth Uhealth* 2013;1(1):e9 [FREE Full text] [doi: [10.2196/mhealth.2709](https://doi.org/10.2196/mhealth.2709)] [Medline: [25100681](https://pubmed.ncbi.nlm.nih.gov/25100681/)]
80. GitHub: where the world builds software [Software repository]. GitHub, Inc. 2021. URL: <https://github.com/> [accessed 2021-12-21]
81. Higgins K, Tutelman P, Chambers C, Witteman H, Barwick M, Corkum P, et al. Availability of researcher-led eHealth tools for pain assessment and management: barriers, facilitators, costs, and design. *Pain Reports* 2018 Sep;3(1):e686. [doi: [10.1097/pr9.0000000000000686](https://doi.org/10.1097/pr9.0000000000000686)]
82. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. *Am J Prev Med* 2016 Dec;51(6):1051-1059. [doi: [10.1016/j.amepre.2016.07.009](https://doi.org/10.1016/j.amepre.2016.07.009)] [Medline: [27659122](https://pubmed.ncbi.nlm.nih.gov/27659122/)]
83. The growing value of digital health. QVIA Institute. 2017. URL: <https://www.iqvia.com/> [accessed 2021-12-21]
84. Anthes E. Pocket psychiatry - mobile mental-health apps have exploded onto the market, but few have been thoroughly tested. *Nature* 2016 Apr 7;532(7597):20-23 [FREE Full text]
85. Ghassemlou S, Marini C, Chemi C, Ranjit YS, Tofighi B. Harmful smartphone applications promoting alcohol and illicit substance use: a review and content analysis in the United States. *Transl Behav Med* 2020 Oct 12;10(5):1233-1242 [FREE Full text] [doi: [10.1093/tbm/ibz135](https://doi.org/10.1093/tbm/ibz135)] [Medline: [33044528](https://pubmed.ncbi.nlm.nih.gov/33044528/)]
86. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. *Mhealth* 2018 Mar;4:6 [FREE Full text] [doi: [10.21037/mhealth.2018.03.02](https://doi.org/10.21037/mhealth.2018.03.02)] [Medline: [29682510](https://pubmed.ncbi.nlm.nih.gov/29682510/)]
87. Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, et al. Validation of the Mobile Application Rating Scale (MARS). *PLoS One* 2020 Nov 2;15(11):e0241480 [FREE Full text] [doi: [10.1371/journal.pone.0241480](https://doi.org/10.1371/journal.pone.0241480)] [Medline: [33137123](https://pubmed.ncbi.nlm.nih.gov/33137123/)]
88. Brzan PP, Rotman E, Pajnikihar M, Klanjsek P. Mobile applications for control and self management of diabetes: a systematic review. *J Med Syst* 2016 Sep;40(9):210. [doi: [10.1007/s10916-016-0564-8](https://doi.org/10.1007/s10916-016-0564-8)] [Medline: [27520615](https://pubmed.ncbi.nlm.nih.gov/27520615/)]
89. Marshall J, Dunstan D, Bartik W. The digital psychiatrist: in search of evidence-based apps for anxiety and depression. *Front Psychiatry* 2019;10:831 [FREE Full text] [doi: [10.3389/fpsy.2019.00831](https://doi.org/10.3389/fpsy.2019.00831)] [Medline: [31803083](https://pubmed.ncbi.nlm.nih.gov/31803083/)]
90. Tofighi B, Chemi C, Ruiz-Valcarcel J, Hein P, Hu L. Smartphone apps targeting alcohol and illicit substance use: systematic search in commercial app stores and critical content analysis. *JMIR Mhealth Uhealth* 2019 Apr 22;7(4):e11831 [FREE Full text] [doi: [10.2196/11831](https://doi.org/10.2196/11831)] [Medline: [31008713](https://pubmed.ncbi.nlm.nih.gov/31008713/)]
91. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. *J Biomed Inform* 2017 Dec;71:31-48. [doi: [10.1016/j.jbi.2017.05.011](https://doi.org/10.1016/j.jbi.2017.05.011)] [Medline: [28536062](https://pubmed.ncbi.nlm.nih.gov/28536062/)]
92. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. *BMJ Open* 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: [10.1136/bmjopen-2016-012447](https://doi.org/10.1136/bmjopen-2016-012447)] [Medline: [27707829](https://pubmed.ncbi.nlm.nih.gov/27707829/)]
93. Schmidt-Kraepelin M, Thiebes S, Sunyaev A. Investigating the relationship between user ratings and gamification - A review of mHealth apps in the apple App Store and Google play store. In: *Proceedings of the Hawaii International Conference on System Sciences*. 2019 Presented at: Hawaii International Conference on System Sciences; 2019; Hawaii p. 1496-1505. [doi: [10.24251/hicss.2019.181](https://doi.org/10.24251/hicss.2019.181)]
94. Foster E, Hawkins A, Simpson E, Adamson AJ. Developing an interactive portion size assessment system (IPSAS) for use with children. *J Hum Nutr Diet* 2014 Jan 19;27 Suppl 1:18-25. [doi: [10.1111/jhn.12127](https://doi.org/10.1111/jhn.12127)] [Medline: [23682796](https://pubmed.ncbi.nlm.nih.gov/23682796/)]
95. Timon CM, Blain RJ, McNulty B, Kehoe L, Evans K, Walton J, et al. The development, validation, and user evaluation of foodbook24: a web-based dietary assessment tool developed for the Irish adult population. *J Med Internet Res* 2017 May 11;19(5):e158 [FREE Full text] [doi: [10.2196/jmir.6407](https://doi.org/10.2196/jmir.6407)] [Medline: [28495662](https://pubmed.ncbi.nlm.nih.gov/28495662/)]
96. Bucci S, Schwannauer M, Berry N. The digital revolution and its impact on mental health care. *Psychol Psychother* 2019 Jun 28;92(2):277-297. [doi: [10.1111/papt.12222](https://doi.org/10.1111/papt.12222)] [Medline: [30924316](https://pubmed.ncbi.nlm.nih.gov/30924316/)]
97. Boumparis N, Schulte MH, Ripper H. Digital mental health for alcohol and substance use disorders. *Curr Treat Options Psych* 2019 Nov 26;6(4):352-366. [doi: [10.1007/s40501-019-00190-y](https://doi.org/10.1007/s40501-019-00190-y)]
98. Martin-Key N, Schei T, Barker E, Spadaro B, Funnell E, Benacek J, et al. The current state and diagnostic accuracy of digital mental health assessment tools for psychiatric disorders: protocol for a systematic review and meta-analysis. *JMIR Res Protoc* 2021 Jan 08;10(1):e25382 [FREE Full text] [doi: [10.2196/25382](https://doi.org/10.2196/25382)] [Medline: [33416508](https://pubmed.ncbi.nlm.nih.gov/33416508/)]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test

AUDIT-C: Alcohol Use Disorders Identification Test–Consumption

DAT: digital assessment tool

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

PROSPERO: International Prospective Register of Systematic Reviews

SBI: screening and brief intervention

Edited by A Mavragani; submitted 21.03.21; peer-reviewed by MDG Pimentel, F Bos, C Reis; comments to author 18.08.21; revised version received 08.11.21; accepted 06.12.21; published 23.03.22.

Please cite as:

Wiemker V, Neufeld M, Bunova A, Danquah I, Ferreira-Borges C, Konigorski S, Rastogi A, Probst C

Digital Assessment Tools Using Animation Features to Quantify Alcohol Consumption: Systematic App Store and Literature Review
J Med Internet Res 2022;24(3):e28927

URL: <https://www.jmir.org/2022/3/e28927>

doi: [10.2196/28927](https://doi.org/10.2196/28927)

PMID: [35319472](https://pubmed.ncbi.nlm.nih.gov/35319472/)

©Veronika Wiemker, Maria Neufeld, Anna Bunova, Ina Danquah, Carina Ferreira-Borges, Stefan Konigorski, Ankit Rastogi, Charlotte Probst. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 23.03.2022. 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.

Review

The Effectiveness of Physical Activity-Promoting Web- and Mobile-Based Distance Weight Loss Interventions on Body Composition in Rehabilitation Settings: Systematic Review, Meta-analysis, and Meta-Regression Analysis

Heli Lahtio^{1,2}, MSc; Aki Rintala^{2,3}, PhD; Jaakko Immonen⁴, MSc; Tuulikki Sjögren¹, PhD

¹Faculty of Sport and Health Sciences, The University of Jyväskylä, Jyväskylä, Finland

²Faculty of Social Services and Health Care, LAB University of Applied Sciences, Lahti, Finland

³Department of Neurosciences, Center for Contextual Psychiatry, KU Leuven, Leuven, Belgium

⁴Faculty of Mathematics and Science, The University of Jyväskylä, Jyväskylä, Finland

Corresponding Author:

Heli Lahtio, MSc

Faculty of Sport and Health Sciences

The University of Jyväskylä

Keskussairaalantie 4

Jyväskylä, 40014

Finland

Phone: 358 14 260 1211

Email: heli.lahtio@gmail.com

Abstract

Background: Overweight and obesity are major problems worldwide, and they lead to an increased risk for several diseases. The use of technology in the treatment of obesity is promising, but in the existing literature, there is considerable uncertainty regarding its efficacy. In this review, we included web- and mobile-based weight loss interventions that were implemented remotely in rehabilitation settings.

Objective: The aim of this systematic review is to study the effectiveness of physical activity-promoting web- and mobile-based distance weight loss interventions in rehabilitation settings on body composition in comparison with control groups that did not use technology.

Methods: Studies were searched from 9 databases. The inclusion criteria were as follows: population: age 18-65 years; intervention: physical activity-promoting web- and mobile-based distance weight loss interventions; comparison: control groups without the use of technology; outcome: changes in BMI, waist circumference, or body fat percentage; study design: randomized controlled trial. The quality of the studies was assessed by 2 researchers. Meta-analysis was performed, and we also conducted a meta-regression analysis to evaluate the factors associated with the changes in body composition outcomes if statistical heterogeneity was observed.

Results: The meta-analysis included 30 studies. The mean quality of the studies was 7 of 13 (SD 1.9; range 3-10). A statistically significant difference was observed in BMI (mean difference [MD] 0.83, 95% CI 0.51-1.15 kg/m²; $P < .001$), waist circumference (MD 2.45, 95% CI 1.83-3.07 cm; $P < .001$), and body fat percentage (MD 1.07%, 95% CI 0.74%-1.41%; $P < .001$) in favor of the weight loss groups using web- or mobile-based interventions. Meta-regression analyses found an association between personal feedback and BMI ($P = .04$), but other factors did not play a role in explaining statistical heterogeneity.

Conclusions: Web- and mobile-based distance weight loss interventions significantly reduced BMI, waist circumference, and body fat percentage. Future studies should focus on the comparability of the intervention content. Future studies are needed to better understand weight loss and identify which components are essential in achieving it.

Trial Registration: PROSPERO CRD42016035831; <https://tinyurl.com/7c93tvd4>

(*J Med Internet Res* 2022;24(3):e25906) doi:[10.2196/25906](https://doi.org/10.2196/25906)

KEYWORDS

technology; weight loss; rehabilitation; overweight; obesity; body mass index; waist circumference; body fat percentage; mobile phone

Introduction

Background

Overweight and obesity are conditions that may influence people's health [1]. In 2016, more than 1.9 billion adults were overweight and 650 million were obese [1]. Overweight is defined as a BMI of ≥ 25 kg/m², and obesity is defined as a BMI of ≥ 30 kg/m² [1]. However, the BMI categories seem to differ based on cultural and racial backgrounds. For example, in Asian populations, the BMI categories are as follows: < 18.5 kg/m² is underweight, 18.5 kg/m² to 23 kg/m² is increasing but acceptable risk, 23 kg/m² to 27.5 kg/m² is increased risk, and ≥ 27.5 kg/m² is high risk [2]. BMI is widely used to define overweight and obesity because it is simple to use, and it is the same for adults of all genders and ages [1]. However, BMI does not consider the distribution of fat, and it does not separate the weight of muscle mass from that of fat mass [3].

Waist circumference has been used to identify abdominal fat distribution [3]. Abdominal fat is metabolically more active than fat in the hips or thighs [3]. However, waist circumference is an imprecise indicator because it does not separate subcutaneous fat from visceral fat [4]. Another limitation is that the cutoff points cannot be applied universally [5]. However, increased waist circumference has been shown to be associated with cardiovascular diseases such as hypertension [5]. Body fat percentage provides valuable information about the distribution of fat [3]. Bioimpedance, dual-energy x-ray absorptiometry (DXA), and skinfold thickness are generally used, but their reliability has been called into question. For example, Sun et al [6] found that bioimpedance was reliable in people with normal body fat range, but it tended to underestimate body fat in people who were thin, overweight, or obese. Bioimpedance measures the electrical properties of body tissues as alternating electric current flows through a human body using parameters such as fat-free mass or total body water [7]. Another example is DXA, which is a 3-component model (1 component of fat mass and 2 components of fat-free mass), but it is not widely used because of its high cost and risk of unnecessary radiation exposure for patients [8].

Several studies have highlighted the role of the combination of physical activity and nutrition in weight loss [9-12]. However, physical activity and nutrition are not the only features affecting weight loss. For example, the gastrointestinal tract, genetics, psychological stress, and medication also affect weight loss [13]. Physical activity has a large role in weight loss in maintaining the amount of fat-free mass and improving the body composition of participants [14]. In addition, it has many health benefits [10,15,16], such as improving the cardiorespiratory system and muscular strength, that are associated with health and functional capacity [17]. It also improves blood pressure and increases insulin sensitivity and fat oxidation [15]. Physical

inactivity increases the risk of several diseases [16,18] as does overweight [3].

Physical activity is an important factor in the prevention and treatment of noncommunicable diseases such as stroke, heart diseases, diabetes, and breast and colon cancer. It is also important in the prevention of risk factors of noncommunicable diseases, such as hypertension. Physical activity can be defined as a movement that has been produced by the skeletal muscles and requires energy expenditure [19]. Physical activity is an integral part of rehabilitation because it has a variety of effects on many organs and functions, such as lowering blood pressure [20]. Physical activity-based rehabilitation has been used, for example, in home-based rehabilitation [20], cardiac rehabilitation [21], and spinal cord injury rehabilitation [22], as well as with various chronic disabilities (eg, musculoskeletal or neurological problems) [23]. Rehabilitation plays a critical role in preventing as well as in minimizing the limitations of functioning that are associated with different conditions [24]. Rehabilitation is needed by anyone with health conditions, not only people with disabilities [24]. In this review, we were interested in weight loss interventions in rehabilitation settings and because of the role of physical activity in rehabilitation [19,25], we included only studies that included physical activity.

The use of technology has increased in the rehabilitation environment in the past decade. For example, technology has shown benefits for rehabilitation focused on cardiac rehabilitation [26,27], nonalcoholic fatty liver disease [28], serious mental illness [29,30], and older adults [31]. These studies focused on weight loss [27-29,31], lifestyle changes [30], and decreasing coronary risk factors [26]. The studies contained, for example, mobile health technology (eg, smartphones and SMS text messages) and social media to increase motivation and facilitate self-monitoring and peer support [29]; a Facebook group-based lifestyle program where participants were able to connect and support each other with regard to healthy eating and exercise goals [30]; an SMS text messaging intervention where the messages provided education on nutrition, exercise, and stress management, as well as improving motivation [28]; videoconferencing sessions consisting of nutrition and exercise sessions [31]; and mailed written materials, a pedometer, and coaching and goal-setting sessions through the telephone [27]. The study by Varnfield et al [26] included smartphones for monitoring health and exercise and delivered motivational and educational materials. It also provided access to a web portal where participants received weekly consultations [26].

Many systematic reviews have studied the effectiveness of technology-based distance weight loss interventions among adults [32-38], but only in a few reviews was the primary outcome related to body composition [32-34]. In other reviews, the primary outcome was weight change [35-38]. There is evidence that technology-based distance weight loss interventions improved weight loss and body composition

outcomes compared with usual care without the use of technology [35-37], no treatment [34] or minimal intervention (information) [33], pamphlets [34], and self-help written materials (eg, dietary guidelines) [35]. There was a variety in the selected study designs in the previous systematic reviews. For instance, most reviews included randomized controlled trials (RCTs) [32-38], but quasi-RCTs [34], single-group intervention studies [36], non-RCTs [36], comparative effectiveness trials [36], retrospective cohort studies [36], and pre-post and quasi-experimental studies [32] were also included. There was variation in the inclusion criteria of the reviews in terms of control groups. Two reviews [34,35] determined what content they would accept for the control groups. They included usual care [34,35], educational materials [34], telephone interventions [34], in-person interventions [34,35], a no-intervention control group [35], or another eHealth intervention [35]. One review [32] had no restrictions about the content of the control group, 2 reviews [33,37] did not mention the content of the control group in the inclusion criteria, and in 2 reviews [36,38], all the studies did not include a control group. There was statistical heterogeneity in most of the reviews [32-37].

Only three previous reviews have used body composition outcomes: BMI [32,34] and waist circumference [33,34]. None of these reviews included all body composition outcomes (BMI, waist circumference, and body fat percentage). Other reviews investigated solely the effectiveness of weight loss (measured in kilograms) [35-37]. Although these systematic reviews have presented positive results, there is also contrary evidence. In body composition outcomes, no difference between the technology user group and control group was found [32]; technology-based interventions achieved smaller weight loss than in-person interventions [34]; or there were inconsistent results [38]. Previous reviews have been heterogeneous in terms of technology. They have described the technology used as eHealth interventions [32,35,36], internet-based interventions [33,37,38], an intervention that used a PC or mobile device [37], or an interaction-enabled computer-based intervention [34]. This review has focused on web- and mobile-based interventions, and the outcome variables were BMI, waist circumference, and body fat percentage.

Only 1 previous systematic review [33] has used meta-regression analysis to investigate the association between personal characteristics and waist circumference. The authors found that baseline waist circumference, gender, and social support were significantly associated with a reduction in waist circumference [33]. However, the meta-regression analysis only focused on waist circumference and studies that included internet-based interventions. Similar meta-regression analyses have not been conducted on other outcomes such as BMI and body fat percentage, which may increase the knowledge that should be considered in future treatment strategies for weight loss interventions. This meta-regression analysis provides a more comprehensive analysis than other previous reviews. In addition, this review discusses how different studies have implemented distance weight loss interventions that use web- and mobile-based intervention.

It has been suggested that technology-based distance health promotion interventions should be theory-based [39] and strengthen self-regulatory skills, which are essential in maintaining behavior change [40]. Widely used theories in technology-based weight loss interventions are cognitive behavioral [41] and social cognitive theories [40]. The cognitive behavioral theory is based on the assumption that all behavior is learned and internal and environmental factors are related to the behavior. It teaches to explore, identify, and analyze dysfunctional patterns of thinking and acting. Strategies related to cognitive behavioral theory include self-monitoring, goal setting, and social support. [41]. According to the social cognitive theory, people learn by observing social interactions and experiences. According to the social cognitive theory, effective strategies are demonstration, modeling, and social support [41].

Objective

The aim of this systematic review is to study the effectiveness of physical activity-promoting web- and mobile-based distance weight loss interventions in rehabilitation settings on body composition in comparison with control groups that do not use technology. We conducted a comprehensive systematic review with meta-analysis of absolute changes in BMI, waist circumference, and body fat percentage. We also conducted a meta-regression analysis to evaluate the factors associated with the changes in body composition outcomes if statistical heterogeneity was observed.

Methods

Data Sources

The following 9 databases were searched from January 2000 to January 2016: PsycINFO, Cochrane Central Register of Controlled Trials, Embase, CINAHL, Ovid MEDLINE, OTseeker, PEDro, Web of Science, and PubMed. An updated search of the following 5 databases was performed from January 1, 2016, to March 31, 2020: PsycINFO, CINAHL, Ovid MEDLINE, PEDro, and PubMed. It was decided to conduct the updated search only in the databases from which we retrieved all the included studies in the first search. Our updated search resulted in 2684 studies, and this number is considered sufficient; it is unlikely that the updated search would have resulted in missing out on studies relevant to our review. In all, 2 information specialists (HL and AR) performed the searches using search terms related to technology and physical activity as well as terms reflecting RCT and clinical trial study designs. The search strategy is presented in [Multimedia Appendix 1](#). A manual search was conducted using the reference lists acquired from the studies found in the search. This systematic review is registered with PROSPERO (CRD42016035831).

Study Selection

With the support of a research team, 1 researcher (HL) was responsible for searching for studies related to the outcomes of this study. The studies were screened by 1 reviewer (HL) using the Population, Intervention, Comparison, Outcomes, and Study Design (PICOS) approach recommended in the PRISMA (Preferred Reporting Items for Systematic Reviews and

Meta-Analyses) checklist [42]. The inclusion criteria were set according to the PICOS framework and were as follows: population: adults aged 18-65 years; intervention: physical activity-promoting web- and mobile-based distance weight loss interventions in rehabilitation settings in the experimental group; comparison: control group without the use of technology; outcome: changes in BMI, waist circumference, or body fat percentage; and study design: RCT. We decided to set the upper age limit as 65 years because Finnish legislation categorizes people aged >65 years as elderly [43]. Rehabilitation is defined as a set of interventions designed to optimize functioning and reduce the disability of individuals who have health conditions [14]. As this review explores the benefits of web- and mobile-based interventions in rehabilitation settings, the interventions in the included studies had to involve participants requiring rehabilitation. We defined web-based interventions as interventions that are performed through the internet and mobile-based interventions as interventions that are performed through mobile devices [44]. Interventions had to be implemented remotely so that the participant was able to use the technological device without the presence of the health care provider. The only exceptions were monthly measurement visits or introduction lessons, seminars, or discussion visits. As outcome variables, we used European BMI categories. However, in different racial groups, BMI categories differ. Studies published in Finnish, English, or Swedish were included. In the reporting of the review, the PRISMA guidelines [42] were adhered to.

Methodological Quality of the Included Studies

The methodological quality of the RCTs was evaluated using the 13-point scale introduced in the guidelines for systematic reviews in the Cochrane Collaboration Back Review Group [45]. Two assessors (HL and AR) from the research group evaluated the studies. Both assessors evaluated the studies independently at first, after which the results were discussed. If necessary, a third reviewer (TS) was consulted to resolve any disagreement. However, a consensus was reached for each study. If a criterion was fulfilled, the domain was considered a *Yes* and counted as a score. If a criterion was not fulfilled, it was considered a *No*. If a criterion was unclear, it was considered a *Don't know*, symbolized by ?. The main quality domains were the following: A1: randomization; B2: concealed treatment allocation; C3: blinding of the patients; C4: blinding of the care providers; C5: blinding of the outcome assessors; D6: dropout rate; D7: analysis of participants in the groups to which they have been assigned; E8: selective outcome reporting; F9: similarity of groups at baseline; F10: similarity or absence of cointerventions; F11: compliance; F12: timing of outcome assessments, and F13: other types of biases. After summing the *Yes* scores, the maximum score obtainable was 13. The quality points were used as a covariate in the meta-regression analysis.

Data Analysis

The meta-analysis consisted of mean difference (MD) analyses for three parameters: BMI, waist circumference, and body fat percentage. A positive MD was considered to favor the experimental group. MD was calculated using groupwise MDs; if these groupwise MDs (and corresponding SDs or SEs) were

reported inadequately, only end point measures were used. Instead of SDs or SEs, some studies reported *P* values or CIs. In this case, the corresponding SDs were calculated from *P* values or CIs, assuming a 2-tailed *t* test. If the median was reported instead of the mean, the median was used to estimate the mean by using the formula $\text{range}/4$ [46]. The authors of the included studies were contacted if the required data were missing or not adequately reported. Responses were received from 11 authors.

All 3 analyses were conducted in 2 phases. First, an ordinary meta-analysis (fitted with restricted maximum likelihood) with no covariates was conducted to estimate the average MD, the heterogeneity among the studies, and the possibility of publication bias. Analyses were conducted using R with a meta-analysis package (metafor; The R Foundation for Statistical Computing) [47]. The level of heterogeneity was measured using the I^2 measure, and its significance was tested using the Cochran Q test. Thresholds for the interpretation of the I^2 measure were as follows: low (0%-40%), moderate (30%-60%), substantial (50%-90%), and considerable (75%-100%) [48]. Publication bias was assessed visually using funnel plots and by using a regression test. The regression test measures the association between the effect sizes and the corresponding SEs [49]. If the funnel plot is symmetrical, publication bias does not exist [48].

Second, the influence of covariates was studied using meta-regression analysis. Forward selection starting with a model without covariates was used. Model fit was measured by corrected Akaike information criterion (AICc): the smaller the AICc, the better the model fit. Models were fitted with maximum likelihood, which enables model comparison. Finally, the model to be chosen (if not the same as in the first phase) was fitted with restricted maximum likelihood to produce unbiased *P* values and CIs. A 3-level meta-analytic model [50] was applied in both phases. Dependence induced by multiarm studies was treated with the method proposed in the study by Gleser and Olkin [51].

Meta-regression analysis was used to explain the statistical heterogeneity. A prior decision was made to investigate the study-level characteristics. Covariates were related to the PICOS framework: population: mean age of the participants, the proportion of men, and analysis of the prevention; intervention: length of the intervention, personal goals, self-reporting, personal feedback, and using theory; comparison: the content of the control group; study design: quality of the studies (for the coding of these variables, see Table 1).

The 9 covariates chosen for the meta-regression analysis were based on the findings from previous systematic reviews. Covariates related to the population were based on mixed findings of previous studies. The results were mixed with regard to weight loss and age [52,53] or gender [37,52,54] of the participants. In previous reviews, too, the results were mixed in studies concerning primary prevention [34], secondary prevention [33,55,56], and tertiary prevention [52,57]. Primary prevention means that a participant is healthy and the prevention is targeted at preventing diseases. In secondary prevention, a participant has symptoms of a disease, and the actions are

targeted to prevent more symptoms or the development of a disease. For example, a participant may have increased blood pressure and the actions are targeted to prevent hypertension. Tertiary prevention means that a participant already has a disease such as hypertension and the actions are targeted to decrease the effect of the disease [58]. Regarding the interventions, the duration of weight loss interventions [52,59] and the best methods to achieve greater weight loss [34,57,60] were considered. Previous systematic reviews have indicated that self-reporting [34,61], intervention personalization [61], in-person feedback, and targeted structured lifestyle coaching are important in facilitating weight loss [34]. Because of these results, personal goals, self-reporting, and personal feedback were chosen as covariates. Previous reviews [33,37,38] have observed that several studies use theories; therefore, using a

theory was also chosen as a covariate. Regarding comparison, we decided to study whether the content of control groups influences the heterogeneity. In this review, the content of the control groups varied greatly. A study by Johns et al [62] found that control groups who received more advice or counseling lost more weight than control groups who received, for example, only pamphlets. We generated five categories that were compared with a wait-list or no intervention (0): (1) usual care, (2) usual care plus minimal guidance (eg, pamphlets), (3) paper instructions, (4) paper instructions plus minimal instructions (eg, introduction lesson); and (5) other intervention concerning weight loss. Finally, we included a covariate reflecting the level of study quality because previous systematic reviews have discussed the variability of the quality of the studies [33,38,57,60,63] and the heterogeneity of the studies [33,55-57].

Table 1. Description and coding of the covariates used in the meta-regression analysis.

Covariate	Description	Coding
Mean age	<ul style="list-style-type: none"> Mean age of the participants 	Years (continuous variable; range 20.0-69.1)
Proportion of men	<ul style="list-style-type: none"> Relative proportion of men from 0% to 100% 	Range 0-1; the scale has been changed
Analysis of prevention	<ul style="list-style-type: none"> Secondary and tertiary prevention were compared with primary prevention 	S: secondary prevention; T: tertiary prevention
Length of the intervention	<ul style="list-style-type: none"> Length of the intervention 	Weeks (continuous variable; range 4-8)
Personal goals	<ul style="list-style-type: none"> If the intervention included setting personal goals for the weight loss intervention 	No (0) or Yes (1; dichotomous variable)
Self-reporting	<ul style="list-style-type: none"> If the intervention included self-reporting of values needed in the study (eg, weight, steps, and diet) 	No (0) or Yes (1; dichotomous variable)
Personal feedback	<ul style="list-style-type: none"> If the participants received personal feedback about their progress (eg, weight loss or physical activity) 	No (0) or Yes (1; dichotomous variable)
Using theory	<ul style="list-style-type: none"> If the intervention used motivational or behavior change theories 	No (0) or Yes (1; dichotomous variable)
Content of the control group	<ul style="list-style-type: none"> 0=wait-list or no-intervention group 1=usual care or content of the control group has not been mentioned 2=usual care plus minimal guidance (eg, pamphlets) 3=paper instructions 4=paper instructions plus minimal instructions (eg, introduction lessons) 5=other intervention concerning weight loss (eg, annual physician appointment and fitness test) 	Six-level factor using the wait-list group or no-intervention group as a reference group (categorical variable)
Quality of the studies	<ul style="list-style-type: none"> The level of quality assessment 	Range 0-13 (continuous variable) [38]

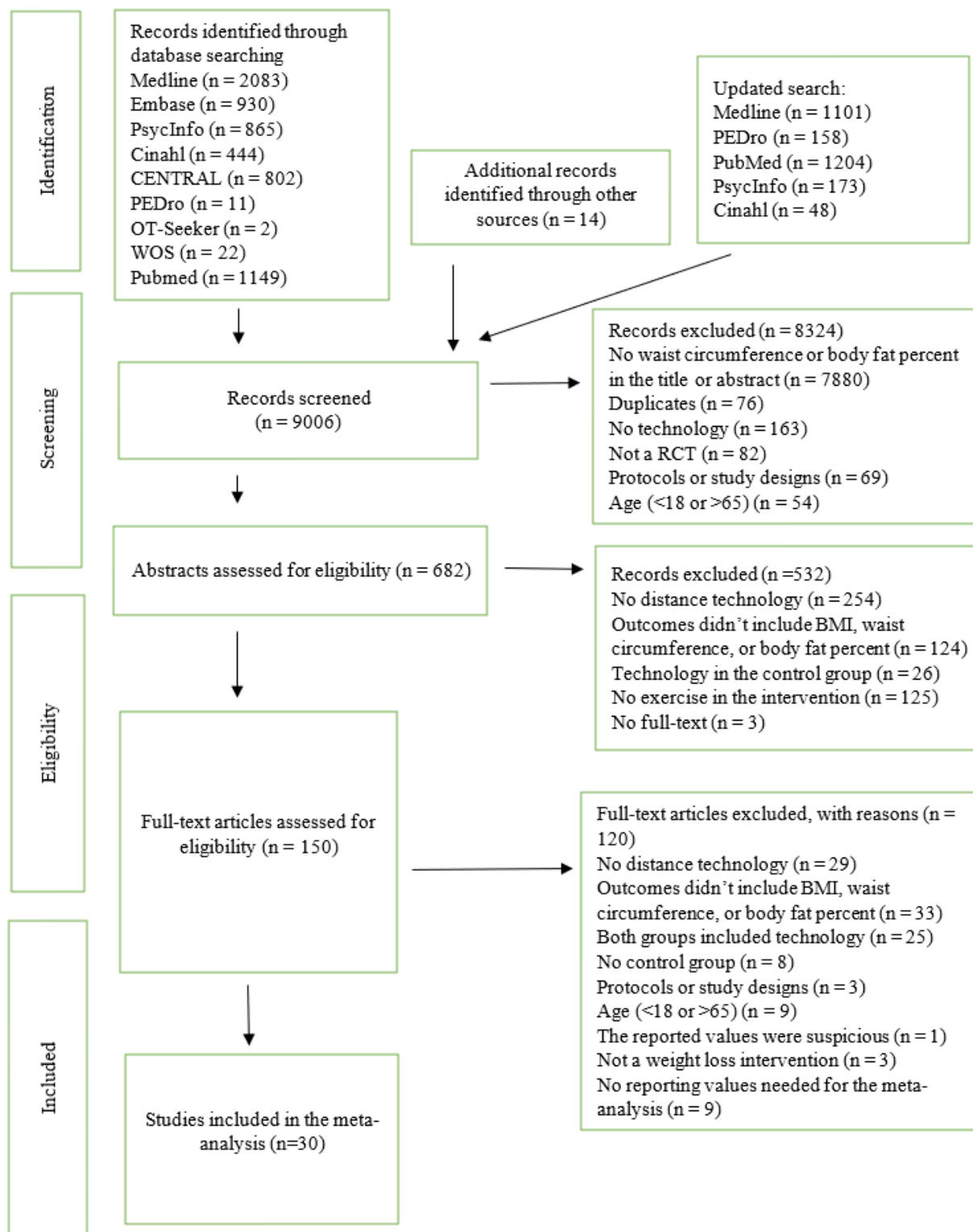
Results

Overview

The search strategy yielded a total of 9006 potentially relevant studies, of which 8976 (99.67%) were excluded, meaning 30 (0.33%) studies met the inclusion criteria and were included in the meta-analysis and meta-regression analysis. A detailed description of the citation screening and selection process is presented in [Figure 1](#).

Among the 30 included studies, the outcome variable was BMI in 19 (63%) studies [64-82], waist circumference in 25 (83%) studies [64,66,68-76,78-91], and body fat percentage in 11 (37%) studies [64,70,72,73,76,78,80,86,89,92,93]. All outcome variables were included in 20% (6/30) of the studies [64,72,74,76,78,80]. In the 11 studies using body fat percentage as an outcome, 7 (64%) used bioimpedance [64,72,73,76,78,89,93], 1 (9%) used DXA [86], and 1 (9%) used skinfold [80], whereas in 2 (18%) studies, the methods were not mentioned [70,92]. A detailed description of the included studies is presented in [Multimedia Appendix 2](#) [43-72].

Figure 1. Flowchart of the study selection.



Methodological Quality of the Studies

The mean methodological quality of the studies was 7 out of 13 (SD 1.9; range 3-10). The quality assessment revealed great variability in the quality of the studies, with the most frequent source of bias being insufficiently reported blinding of the participants or care provider. There was also insufficient reporting on compliance with the intervention, avoidance of cointerventions, dropout descriptions, and analysis of

participants in the assigned groups. The methodological quality of the studies is presented in [Multimedia Appendix 3 \[64-93\]](#).

Description of the Participants

A total of 6103 participants were included in the studies, and the mean percentage of men was 42% (SD 7.7%). The mean age of the participants was 40.2 (SD 17.1) years. In a study by Sakane et al [79], the average age was not reported, but the age range of the participants was used. Therefore, the average age

was imputed from the average age of Japanese people aged 20-65 years [94], weighted with the proportions of men and women in the study. A similar imputation was made by Anderson et al [65], except only the lower bound (18 years) was reported; the average age of American people aged 18-84 years [95], weighted with the proportions of men and women, was used. The number of participants in the experimental groups was 3624 (mean 88.4, SD 98.4), and the mean percentage of men was 38.8% (SD 31.2%). The number of participants in the control groups was 2490 (mean 80.3, SD 106.8), and the mean percentage of men was 41.1% (SD 31.8%). A detailed description of the participants is presented in [Multimedia Appendix 2](#).

Description of the Interventions

The mean duration of the interventions was 30.4 weeks (SD 17.9; range 4 weeks to 2 years). In the included studies, weight loss was set as the main aim of the study in 60% (18/30) of the studies [64,66,68,71-74,76-82,85,86,89,91]. Of the 30 included studies, 3 (10%) aimed to increase the level of physical activity [70,88,92], whereas 2 (7%) focused on both increasing the participants' physical activity and decreasing their weight [84,93]. Of the 30 included studies, 6 (20%) aimed at more specific interventions such as disease management in diabetes [65], lifestyle modification [67], reducing blood pressure [69], health management [83], reducing cardiovascular risk and weight loss [90], and improving quality of life [87]. The included studies were implemented in rehabilitation settings. The instructions for the intervention were given by a health care professional (eg, general practitioner or physician), or the studies were performed in the field of health science (eg, a department of health and physical activity). A primary outcome was weight loss in 37% (11/30) of the studies [64,66,67,71,72,74,76,79,80,82,86], both weight loss and physical activity in 33% (10/30) of the studies [68,73,77,78,81,84,89-91,93], and physical activity in 10% (3/30) of the studies [70,88,92]. Weight change was a secondary outcome in 7% (2/30) of the studies [65,87], and weight loss without definition was the primary or secondary outcome in 10% (3/30) of the studies [69,83,85]. A detailed description of the interventions is presented in [Multimedia Appendix 2](#).

The included studies used mainly mobile- or web-based interventions. Of the 30 included studies, 5 (17%) [65,71,85,87,90] used only mobile phones and 3 (10%) [70,79,83] were only web-based. Both mobile- and web-based interventions were included in 30% (9/30) of the studies [66,69,72,74,75,77,78,80,82]. Altogether, a mobile-based intervention was included in 73% (22/30) of the studies [65,66,68,69,71-75,77,78,80,82,84-91,93], whereas a web-based intervention was included in 60% (18/30) of the studies [64,66,68-70,72,74-80,82,83,87,91,92]. Most of the studies used combinations of technological devices, such as email in 17% (5/30) of the studies [66,68,75,82,91], pedometers in 37% (11/30) of the studies [64,69,73,76,77,84,86,88,89,91,93], and DVDs in 10% (3/30) of the studies [64,69,76]. Other technologies used were a step counter [66], an activity tracker [67], an activity-measuring device [78], digital scales [84], video

clips [74], and an armband [80]. Of the 30 studies, 4 (13%) combined a web-based intervention with a step counter [66] or pedometers [64,76,91], 9 (30%) combined a mobile-based intervention with an activity tracker [67] or pedometers [69,73,84,86,88,89,91,93], and 2 (7%) used both web- and mobile-based interventions with a pedometer [77] or an activity-measuring device [78].

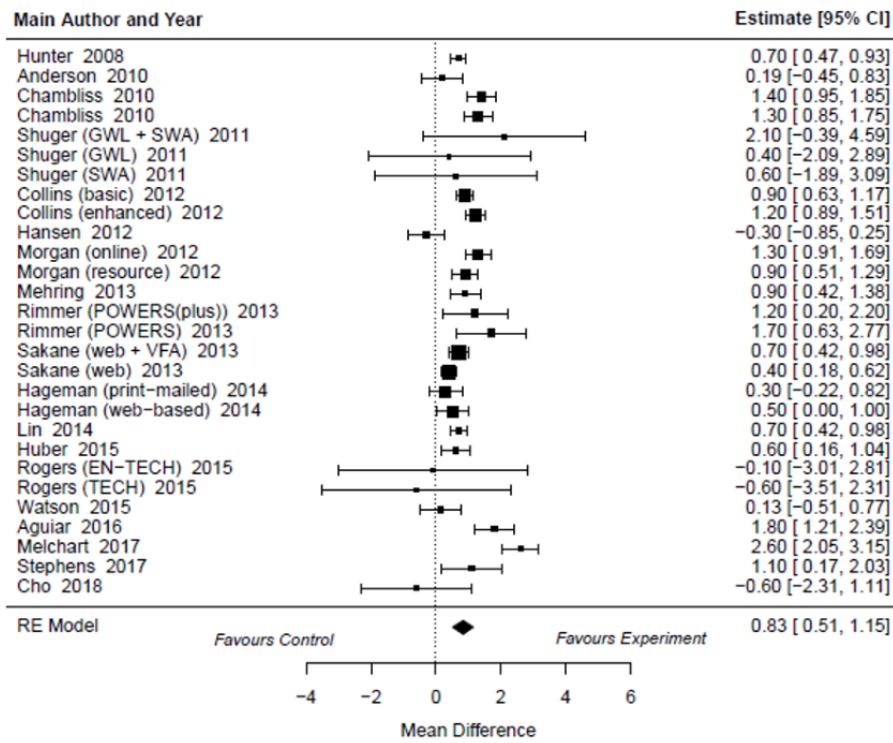
There was a large variation in the frequency of technology use. The participants were contacted by telephone every week [77], every other week [64,71,88,90], every 4-6 weeks [87], or 4 and 8 weeks after baseline measurements [72]. In the study by Rimmer et al [77], telephone calls were reduced from weekly calls to monthly calls, and in the study by Anderson et al [65], call frequency depended on the risk stratification of the participants. There was also variation in the guidance regarding the frequency of using websites, smartphone apps, emails, and accelerometers as a technology. In the studies using websites and smartphone apps, guidance for logging in varied from daily use [66,69,81], use at least once a week [74], use 3-4 times per week [76,92], and use at least five times per week [72]. Of the 30 studies, 5 (17%) did not report the frequency of using a website [68,78,80,83,91]. The frequency of emails was weekly in 7% (2/30) of the studies [68,76]. In the studies using pedometers and accelerometers, guidance for wearing them varied from daily use [69,73,84,86,88] to 2 days a week [64,76,80,89,92]. In the study by Rimmer et al [77], the frequency was not reported.

There was variation in the content of the control groups. The control group included usual care in 27% (8/30) of the studies [65,74,82,84,88,90,92,93], a wait-list in 20% (6/30) of the studies [66,68,76,77,83,85], and minimal treatments such as brochures [67,69,71,86,87,90,91] or an annual physician's examination [72,87] in other studies. Group meetings [78,86], introductory sessions [69], and mailed feedback about the baseline assessment [89] were also used as content in the control groups. Of the 30 studies, 9 (30%) [66,68,69,76,78-80,86,91] included >1 intervention group, all of which were compared with the same control group.

The Effectiveness of Web- and Mobile-Based Distance Weight Loss Interventions on BMI Compared With Control Groups Without the Use of Technology

Web- and mobile-based distance weight loss interventions were 0.83 units more beneficial to BMI than the control groups without the use of technology (MD 0.83, 95% CI 0.51-1.15 kg/m²; $P < .001$; [Figure 2](#)). The studies were considerably heterogeneous ($I^2 = 90\%$; $P < .001$). On the basis of the regression test, there were no signs of publication bias ($P = .61$). According to the AICc, studies using personal feedback have an impact on BMI reduction compared with the included studies that did not have personal feedback (0.32, 95% CI 0.02-0.62; $P = .04$). However, other covariates did not have any effect on the reduction in BMI. Detailed information about the covariates of BMI is presented in [Multimedia Appendix 4](#), and the analysis of BMI is presented in [Multimedia Appendix 5](#) [43-61].

Figure 2. Forest plots describing the effectiveness of web- and mobile-based weight loss intervention on BMI compared with control groups without the use of technology [64,70,72,73,76,78,80,86,89,92,93]. EN-TECH: enhanced technology-based system; GWL: group-based behavioral weight loss program; GWL + SWA: group-based behavioral weight loss program + The SenseWear Armband; RE: random effect; SWA: The SenseWear Armband; TECH: technology-based system.

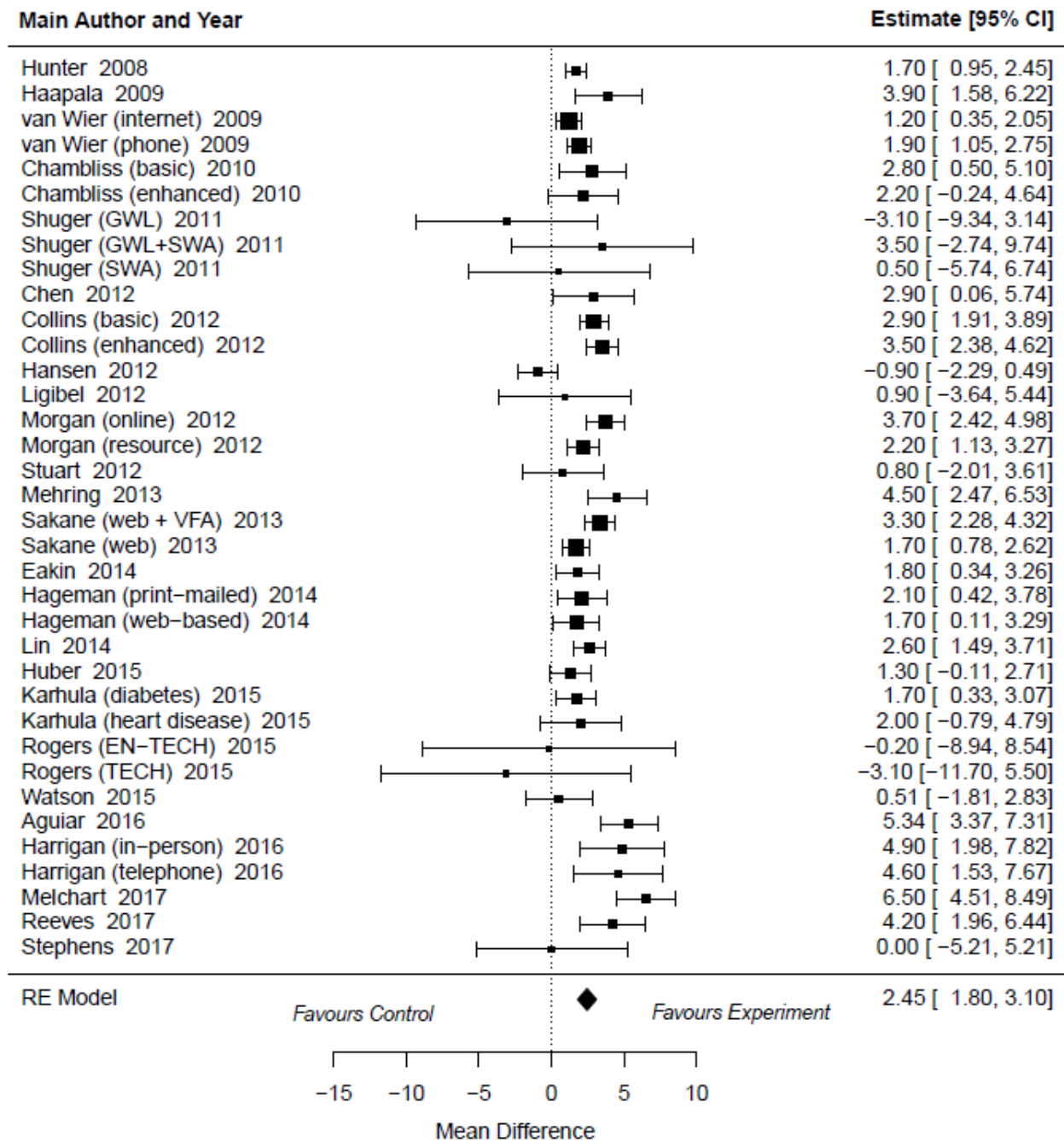


The Effectiveness of Web- and Mobile-Based Distance Weight Loss Interventions on Waist Circumference Compared With Control Groups Without the Use of Technology

Web- and mobile-based distance weight loss interventions were 2.45 units more beneficial to waist circumference than the

control groups without the use of technology (MD 2.45, 95% CI 1.83-3.07 cm; $P < .001$; Figure 3). Substantial heterogeneity was observed ($I^2 = 78\%$; $P < .001$). On the basis of the regression test, there were no signs of publication bias ($P = .73$). According to the AICc, none of the covariates explained the variability in waist circumference reduction (Multimedia Appendices 4 and 6 [43,45,47-55,57-70]).

Figure 3. Forest plots describing the effectiveness of web- and mobile-based weight loss intervention on waist circumference compared with control groups without the use of technology [64,66,68-76,78-91]. EN-TECH: enhanced technology-based system; GWL: group-based behavioral weight loss program; GWL + SWA: group-based behavioral weight loss program + The SenseWear Armband; RE: random effect; SWA The SenseWear Armband; TECH: technology-based system; VFA: visceral fat measurement group.

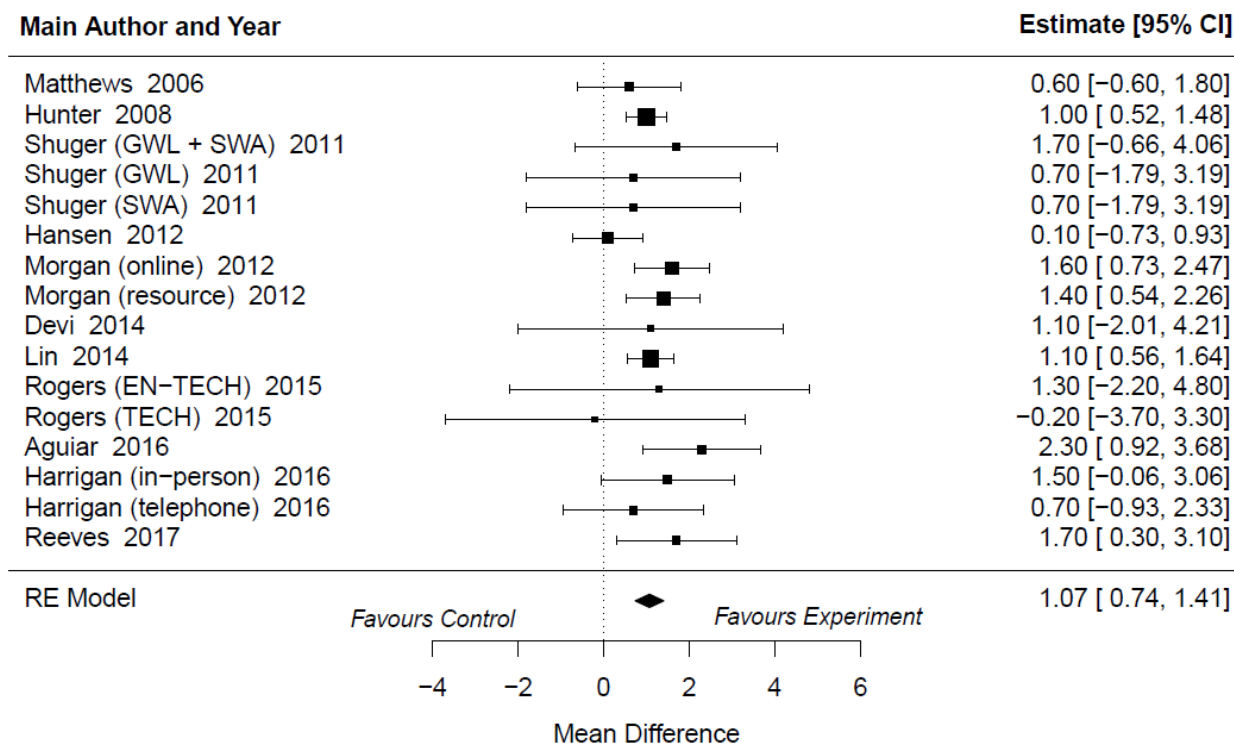


The Effectiveness of Web- and Mobile-Based Distance Weight Loss Interventions on Body Fat Percentage Compared With Control Groups Without the Use of Technology

Web- and mobile-based distance weight loss interventions were 1.07 units more beneficial to body fat percentage than the control groups without the use of technology (MD 1.07%, 95% CI

0.74%-1.41%; $P < .001$; Figure 4). The heterogeneity was low ($I^2 = 18\%$; $P = .54$). On the basis of the regression test, there were no signs of publication bias ($P = .81$). According to the AICc, none of the covariates explained the variability in body fat percentage. However, the quality of the studies indicated that it might have an impact on the reduction in body fat percentage, but it was not statistically significant (estimate 0.12, 95% CI -0.01 to 0.25; $P = .06$; Multimedia Appendices 4 and 7 [43,49,51,52,55,57,59,65,68,71,72]).

Figure 4. Forest plots describing the effectiveness of web- and mobile-based weight loss intervention on body fat percentage compared with control groups without the use of technology [64,70,72,73,76,78,80,86,89,92,93]. EN-TECH: enhanced technology-based system; GWL: group-based behavioral weight loss program; GWL + SWA: group-based behavioral weight loss program + The SenseWear Armband; RE: random effect; SWA: The SenseWear Armband; TECH: technology-based system.



Discussion

Principal Findings

This systematic review and meta-analysis indicated a statistically significant difference in BMI, waist circumference, and body fat percentage in favor of the physical activity-promoting web- and mobile-based distance weight loss experimental groups when compared with the control groups without the use of technology. Previous systematic reviews have also observed that weight loss [32,33,38] and waist circumference reduction [38] were significantly greater in technology-based experimental groups when they were compared with usual care without the use of technology or with minimal interventions (eg, pamphlets). Although in our meta-analysis, the experimental groups achieved greater reduction in BMI, waist circumference, and body fat percentage, the reductions in all outcome variables were relatively small and not clinically meaningful. Previous systematic reviews have indicated similar results [32,96]. They found that although the results were statistically significant, they were not clinically meaningful.

Despite using a vast number of different covariates in our meta-regression analysis, only personal feedback showed a trend toward an effect on the variability of the body composition outcome measurements. Our meta-regression analysis did not indicate that the level of study quality affected the findings, although it was close to statistical significance. This is reassuring for researchers who are conducting RCTs to investigate weight loss. Variability was also considered large in all meta-regression analyses; therefore, a firm scientific conclusion is challenging to draw. The review by Seo et al [33] found that mean waist

circumference at baseline, proportion of male participants, and social support were related to significant waist circumference changes. Other intervention content-related covariates were not significant in the reduction in waist circumference. In the review, the mean age of the participants, existing diseases, and the status of general obesity also had a significant effect on waist circumference reduction [33]. Despite important elements found in previous studies (self-monitoring [8,61], in-person feedback [34], targeted and structured lifestyle coaching [34,61], program use [61], and social support [61]), only personal feedback and the quality of the studies were related to greater changes in body composition in our findings. Khaylis et al [97] qualitatively studied components that are effective in facilitating weight loss. According to the researchers, self-monitoring, counselor feedback and communication, social support, structured programs, and individually tailored programs are effective in technology-based weight loss interventions [97]. In this systematic review, self-reporting, personal goals, and personal feedback were used as covariates, but the study did not find an effect on body composition.

In this review, the interventions of the experimental groups were web- or mobile-based. However, only 27% (8/30) of the studies [65,70,71,79,83,85,87,90] used solely web- or mobile-based technology, whereas the other studies used multiple technologies in their interventions combining web- or mobile-based technology with, for example, a pedometer [88], or a DVD and pedometer [76]. In the included studies, there were none in which the experimental group and control group had similar interventions, with the only difference being that the experimental group used technology and the control group did not. In the experimental group, the content of the

intervention might have included several methods, such as telephone calls, self-reporting, personal feedback, and pedometers, whereas the control group had only brochures. Therefore, it is difficult to identify the real effects of technology-based distance weight loss interventions on body composition. In addition, several other factors, such as those related to motivation and commitment to the intervention, could have affected the results. However, in rehabilitation, communication between the participant and health care provider is essential. Although we recognized that the level of communication and social support may complicate investigating the effectiveness of the technology itself, in rehabilitation settings this is unavoidable. Therefore, we included motivation with personal feedback and personal goals as a factor in the meta-regression analysis because this has been found to be a contributing factor in weight loss interventions [97-99]. According to our findings, these covariates did not influence the results of this study and only personal feedback had a statistically significant result in decreasing BMI.

This review provides insight into how weight loss interventions can be implemented in rehabilitation settings, both with and without the use of technology. We explored the benefits of the web- and mobile-based weight loss interventions that were mainly implemented remotely and promoted physical activity. The interventions included, at minimum, access to a basic web-based program [68] or telephone calls [65]. The most intensive intervention included, for example, a meeting with the health educator, individual reports of measurements, a 2-hour group seminar, basic guidance for healthy eating and physical activity, self-monitoring of daily food intake and physical activity, weekly individual feedback reports on diaries, step counters, monthly email newsletters, telephone consultations, and monthly clinic visits with discussions about behavioral strategies [66]. There was also variation in the content of the control groups. Usual care and wait-list were most commonly used. However, pamphlets, annual physician appointments, self-monitoring by means of paper diaries, and weekly group meetings were also examples of the content of control groups. There are various ways to implement weight loss interventions that use technology. It is not known which features are imperative for achieving weight loss [35], but this review provides a general view of the features that were used in previous studies as well as the minimum and maximum features that were used in previous studies.

A challenge of technology-based interventions might be how to motivate participants to use technology enough. For example, in the study by Hansen et al [70], 71% of the participants did not sign into the website at all, and only 2% signed in several times over a 6-month period. Haapala et al [85] found that the frequency of using the website varied from 3 to 8 times per week. According to the investigators, the participants who achieved a weight loss of >5% reported more log-ins to the website than the participants who achieved a weight loss of <5%. To achieve results using weight loss interventions, it would be important to get participants to engage in the study. In future studies, this could be a covariate because it is an important element of weight loss interventions. Future studies should more

accurately report the adherence of participants to the intervention.

Outcome variables have varied in previous reviews concerning body composition changes. Therefore, the clinical significance of the reduction in BMI, waist circumference, or body fat percentage is difficult to determine. There are also certain challenges in these outcomes. BMI has been used to determine obesity. However, it is a poor indicator of body fat percentage, and it does not capture the location of body fat [100]. A clinically significant reduction in waist circumference has not been determined in previous studies. Han et al [101] found that a 5-10 cm reduction in waist circumference was a realistic goal with great health benefits for White women who were overweight. However, this study was published in 1997, and it is out of date. Future studies are needed to determine what a beneficial reduction in waist circumference is in terms of, for example, the prevention of obesity-related illnesses. Body fat percentage can be measured in multiple ways. In this review, 23% (7/30) of the studies used bioimpedance [64,72,73,76,78,89,93], 3% (1/30) used DXA [86], and 3% (1/30) used skinfold [80]. In 7% (2/30) of the studies, the methods were not mentioned [70,92]. Benito et al [8] compared DXA, bioimpedance, and skinfold measurements in adults who were overweight. According to them, skinfold seems to underestimate the values of body fat percentage. Therefore, DXA and bioimpedance can be better tools for measuring it [8]. Liao et al [102] studied the accuracy and agreement of DXA and bioimpedance results. They found moderate to high correlations between these 2 measurements in estimating total and segmental lean body mass, fat mass, and body fat percentage.

Obesity is a multidimensional phenomenon, and although obesity is a widely studied condition, we still do not understand all aspects of it. Strong misconceptions such as laziness and lack of self-discipline are still related to obesity in health care settings, workplaces and education, and public literature. In addition, people with obesity feel a strong stigma, which can affect their physical and psychological well-being [20]. It is important to understand comprehensively the phenomenon of obesity so that in rehabilitation settings we can support the well-being of people with obesity without stigmatization.

Strengths and Limitations

The strength of this systematic review is its carefully conducted statistical analyses with its strict PICOS criteria and the use of only RCTs. Meta-regression analysis with several covariates was used to explain the heterogeneity of the studies. To the best of our knowledge, this is the first systematic review to use a comprehensive analysis to identify possible statistical heterogeneity while investigating the effectiveness of web- and mobile-based distance weight loss interventions in rehabilitation settings on body composition. A limitation of this systematic review was that the included studies were very heterogeneous. An explanation for the clinical heterogeneity might be the variability in the content of the interventions.

Despite several covariates used in the meta-regression analysis, only personal feedback showed a statistically significant association, and the quality of the studies showed a

nonsignificant association with the reduction in waist circumference and body fat percentage. In addition, there was variation in the widths of the CIs, and all the estimates were small. It is also important to note that no multiple testing corrections were made. In addition, there was low statistical power. Therefore, any conclusions drawn from the results of the meta-regression analysis must be treated with caution. According to our findings, the sources of the heterogeneity remain unclear. Some other factors that were not investigated in this study could probably explain the heterogeneity. For example, previous systematic reviews observed associations between social support and waist circumference reduction [33] and between program use and weight loss [61].

The second limitation is that only 1 researcher screened the studies. However, the whole study group supported the study screening, and in unclear cases, the study group was consulted. In the final search, 30 studies met the inclusion criteria; therefore, the review provides a comprehensive view of the topic in question. Third, this review focuses on web- and mobile-based weight loss rehabilitation interventions, which were mostly implemented remotely by a health care professional. Because of this, we cannot generalize the findings of this review to all persons and weight loss interventions. It is possible that other factors in the rehabilitation context, such as social support provided by a health care professional, may have affected the findings. Despite these limitations, we believe that the important findings of this review provide a unique overview of web- and mobile-based distance weight loss interventions to clinicians and researchers in rehabilitation settings.

In this review, 87% (26/30) of the studies were from high-income western countries. Only 13% (4/30) of the studies were from Asia [67,73,79,83]; the rest were from the United States, Australia, or Europe. Because of the high representation of high-income western countries, we were not able to perform a meta-regression analysis for identifying differences across continents. Therefore, the findings of our results cannot be generalized worldwide and more future studies are recommended to investigate the potential of, and possible differences in, the use of technology to support weight loss and to reduce obesity in different cultures.

Future Recommendations

Future studies should focus on the comparability of the intervention content. To study weight loss more specifically, more attention should be paid to the comparability of RCTs and control group content. Studies should focus on interventions where the only difference between the experimental and control groups would be the use of technology; in other respects, the content of the groups would be similar. With this study design, it would be possible to determine the effect of technology in a weight loss intervention. In our review, the content of the interventions in the included studies was heterogeneous, and this may influence the generalizability of our findings. However, this review provides a glance at technology-based interventions

in weight loss rehabilitation. Future studies should use the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [103] to align the terminology for the use of technology and, in turn, increase the quality of future systematic reviews.

In our meta-regression analysis, we considered different aspects of the intervention content to explain the statistical heterogeneity. A study concerning the remote rehabilitation of cardiac rehabilitees' by means of technology [104] suggested that interventions are encouraged to include social participation, such as peer group discussion and personalized feedback. In our review, we were unable to include such covariates to describe the role of communication because only 7% (2/30) of the studies [56,78] used 2-way communication and only 17% (5/30) of the studies included peer support [68,70,74,79,81] through technology. Other studies included a 1-way communication approach. Although we could not use communication as a covariate, we used personal feedback as a covariate. It had an impact on BMI but not on waist circumference or body fat percentage. Future studies should consider communication as an essential tool in supporting the weight loss of participants.

Future studies are also needed to better understand weight loss and identify which components are essential in achieving it. Several covariates did not explain the variability of the results, which may indicate that web- and mobile-based distance interventions in rehabilitation settings in the home environment may be more complex to study and require a more personalized approach. Therefore, in the future, it would be recommended to study the meanings and perceptions of participants on the use of technology in distance rehabilitation settings. As in the cardiac rehabilitees' technology study [104], more personalized approaches should be used in weight loss interventions. Rehabilitation is a complex trust-building process in which, for example, cardiac rehabilitees desire personalized and individualized counseling to maintain motivation [105]. Although weight loss has been widely studied, we still do not understand its phenomena. An aspect that is closely related to personalization is the understanding of obesity-related stigma, which affects the well-being of people with obesity [20]. Future studies should investigate the experiences of stigma and increase the awareness of obesity-related factors as well as our understanding of the stigma for people with obesity.

Conclusions

Web- and mobile-based distance weight loss interventions might be more effective than weight loss interventions without the use of technology. However, the changes in body composition outcomes in this review were not clinically meaningful, and statistical and clinical heterogeneity were present. Future studies are needed to better understand weight loss and identify what components are essential in achieving it.

Acknowledgments

This work was supported by the Social Insurance Institution of Finland (grant number 31/26/2014).

Authors' Contributions

AR was responsible for the quality assessment and for writing the paper. JI conducted the data analysis, and TS carried out the research.

Conflicts of Interest

None declared.

Multimedia Appendix 1

An example of the search strategy.

[DOC File , 47 KB - [jmir_v24i3e25906_app1.doc](#)]

Multimedia Appendix 2

Description of the included studies.

[DOC File , 120 KB - [jmir_v24i3e25906_app2.doc](#)]

Multimedia Appendix 3

Quality points of the included studies.

[DOC File , 90 KB - [jmir_v24i3e25906_app3.doc](#)]

Multimedia Appendix 4

Results of the meta-regression analysis of BMI, waist circumference, and body fat percentage.

[DOC File , 63 KB - [jmir_v24i3e25906_app4.doc](#)]

Multimedia Appendix 5

Main results of the studies included in the meta-analysis of BMI.

[DOC File , 154 KB - [jmir_v24i3e25906_app5.doc](#)]

Multimedia Appendix 6

Main results of the studies included in the meta-analysis of waist circumference.

[DOC File , 183 KB - [jmir_v24i3e25906_app6.doc](#)]

Multimedia Appendix 7

Main results of the studies included in the meta-analysis of body fat percentage.

[DOC File , 105 KB - [jmir_v24i3e25906_app7.doc](#)]

References

1. WHO definition of Health. World Health Organization. URL: <https://www.who.int/about/governance/constitution> [accessed 2016-02-20]
2. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. *Lancet* 2004;363(9403):157-163. [doi: [10.1016/s0140-6736\(03\)15268-3](https://doi.org/10.1016/s0140-6736(03)15268-3)] [Medline: [14726171](#)]
3. World Health Organization. Obesity: preventing and managing the global epidemic: report of a WHO consultation. Geneva: World Health Organization; 2000.
4. Fox CS, Massaro JM, Hoffmann U, Pou KM, Maurovich-Horvat P, Liu C, et al. Abdominal visceral and subcutaneous adipose tissue compartments: association with metabolic risk factors in the Framingham Heart Study. *Circulation* 2007;116(1):39-48. [doi: [10.1161/CIRCULATIONAHA.106.675355](https://doi.org/10.1161/CIRCULATIONAHA.106.675355)] [Medline: [17576866](#)]
5. Nutrition and Food Safety. Waist circumference and waist-hip ratio: report of a WHO expert consultation. Geneva: World Health Organization; 2008.
6. Sun G, French CR, Martin GR, Younghusband B, Green RC, Xie YG, et al. Comparison of multifrequency bioelectrical impedance analysis with dual-energy X-ray absorptiometry for assessment of percentage body fat in a large, healthy population. *Am J Clin Nutr* 2005;81(1):74-78. [doi: [10.1093/ajcn/81.1.74](https://doi.org/10.1093/ajcn/81.1.74)] [Medline: [15640463](#)]
7. Marra M, Sammarco R, De Lorenzo A, Iellamo F, Siervo M, Pietrobelli A, et al. Assessment of body composition in health and disease using bioelectrical impedance analysis (BIA) and dual energy X-ray absorptiometry (DXA): a critical overview. *Contrast Media Mol Imaging* 2019;2019:3548284 [FREE Full text] [doi: [10.1155/2019/3548284](https://doi.org/10.1155/2019/3548284)] [Medline: [31275083](#)]

8. Benito PJ, Gómez-Candela C, Cabañas MD, Szendrei B, Castro EA. Comparison between different methods for measuring body fat after a weight loss program. *Rev Bras Med Esporte* 2019;25(6):474-479. [doi: [10.1590/1517-869220192506149743](https://doi.org/10.1590/1517-869220192506149743)]
9. Ramage S, Farmer A, Eccles KA, McCargar L. Healthy strategies for successful weight loss and weight maintenance: a systematic review. *Appl Physiol Nutr Metab* 2014;39(1):1-20. [doi: [10.1139/apnm-2013-0026](https://doi.org/10.1139/apnm-2013-0026)] [Medline: [24383502](https://pubmed.ncbi.nlm.nih.gov/24383502/)]
10. Jakicic JM, Rogers RJ, Davis KK, Collins KA. Role of physical activity and exercise in treating patients with overweight and obesity. *Clin Chem* 2018;64(1):99-107. [doi: [10.1373/clinchem.2017.272443](https://doi.org/10.1373/clinchem.2017.272443)] [Medline: [29158251](https://pubmed.ncbi.nlm.nih.gov/29158251/)]
11. Chin SH, Kahathuduwa CN, Binks M. Physical activity and obesity: what we know and what we need to know. *Obes Rev* 2016;17(12):1226-1244. [doi: [10.1111/obr.12460](https://doi.org/10.1111/obr.12460)] [Medline: [27743411](https://pubmed.ncbi.nlm.nih.gov/27743411/)]
12. Nascimento SL, Pudwell J, Surita FG, Adamo KB, Smith GN. The effect of physical exercise strategies on weight loss in postpartum women: a systematic review and meta-analysis. *Int J Obes (Lond)* 2014;38(5):626-635. [doi: [10.1038/ijo.2013.183](https://doi.org/10.1038/ijo.2013.183)] [Medline: [24048142](https://pubmed.ncbi.nlm.nih.gov/24048142/)]
13. Rubino F, Puhl RM, Cummings DE, Eckel RH, Ryan DH, Mechanick JI, et al. Joint international consensus statement for ending stigma of obesity. *Nat Med* 2020;26(4):485-497. [doi: [10.1038/s41591-020-0803-x](https://doi.org/10.1038/s41591-020-0803-x)]
14. Hernández-Reyes A, Cámara-Martos F, Molina-Luque R, Romero-Saldaña M, Molina-Recio G, Moreno-Rojas R. Changes in body composition with a hypocaloric diet combined with sedentary, moderate and high-intense physical activity: a randomized controlled trial. *BMC Womens Health* 2019;19(1):167 [FREE Full text] [doi: [10.1186/s12905-019-0864-5](https://doi.org/10.1186/s12905-019-0864-5)] [Medline: [31882009](https://pubmed.ncbi.nlm.nih.gov/31882009/)]
15. Maurer S. Weight management in primary care. *Visc Med* 2016;32(5):342-346. [doi: [10.1159/000450667](https://doi.org/10.1159/000450667)] [Medline: [27921046](https://pubmed.ncbi.nlm.nih.gov/27921046/)]
16. Physical activity. World Health Organization. 2020. URL: <https://www.who.int/news-room/fact-sheets/detail/physical-activity> [accessed 2018-02-23]
17. Strasser B. Physical activity in obesity and metabolic syndrome. *Ann N Y Acad Sci* 2013;1281(1):141-159 [FREE Full text] [doi: [10.1111/j.1749-6632.2012.06785.x](https://doi.org/10.1111/j.1749-6632.2012.06785.x)] [Medline: [23167451](https://pubmed.ncbi.nlm.nih.gov/23167451/)]
18. Booth FW, Roberts CK, Laye MJ. Lack of exercise is a major cause of chronic diseases. *Compr Physiol* 2012;2(2):1143-1211 [FREE Full text] [doi: [10.1002/cphy.c110025](https://doi.org/10.1002/cphy.c110025)] [Medline: [23798298](https://pubmed.ncbi.nlm.nih.gov/23798298/)]
19. Global action plan on physical activity 2018-2030: more active people for a healthier world. World Health Organization. Geneva: World Health Organization; 2018. URL: <https://apps.who.int/iris/bitstream/handle/10665/272722/9789241514187-eng.pdf?sequence=1&isAllowed=y> [accessed 2021-02-22]
20. Loellgen H, Zupet P, Bachl N, Debruyne A. Physical activity, exercise prescription for health and home-based rehabilitation. *Sustainability* 2020;12(24):10230. [doi: [10.3390/su122410230](https://doi.org/10.3390/su122410230)]
21. Thomas RJ, Beatty AL, Beckie TM, Brewer LC, Brown TM, Forman DE, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. *Circulation* 2019;140(1):e69-e89. [doi: [10.1161/cir.0000000000000663](https://doi.org/10.1161/cir.0000000000000663)] [Medline: [31082266](https://pubmed.ncbi.nlm.nih.gov/31082266/)]
22. Jones ML, Harness E, Denison P, Tefertiller C, Evans N, Larson CA. Activity-based therapies in spinal cord injury: clinical focus and empirical evidence in three independent programs. *Top Spinal Cord Inj Rehabil* 2012;18(1):34-42. [doi: [10.1310/sci1801-34](https://doi.org/10.1310/sci1801-34)] [Medline: [23459641](https://pubmed.ncbi.nlm.nih.gov/23459641/)]
23. Røe C, Preede L, Dalen H, Bautz-Holter E, Nyquist A, Sandvik L, et al. Does adapted physical activity-based rehabilitation improve mental and physical functioning? A randomized trial. *Eur J Phys Rehabil Med* 2018;54(3):419-427 [FREE Full text] [doi: [10.23736/S1973-9087.16.04189-7](https://doi.org/10.23736/S1973-9087.16.04189-7)] [Medline: [27171538](https://pubmed.ncbi.nlm.nih.gov/27171538/)]
24. Rehabilitation: key for health in the 21st century. World Health Organization. 2017. URL: https://www.who.int/docs/default-source/documents/health-topics/rehabilitation/call-for-action/keyforhealth21stcentury.pdf?sfvrsn=43cebb7_5 [accessed 2021-02-22]
25. Frontera WR. Physical activity and rehabilitation in elderly. In: Masiero S, Carraro U, editors. *Rehabilitation medicine for elderly patients*. Cham: Springer; 2018:3-13.
26. Varnfield M, Karunanithi M, Lee CK, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. *Heart* 2014;100(22):1770-1779 [FREE Full text] [doi: [10.1136/heartjnl-2014-305783](https://doi.org/10.1136/heartjnl-2014-305783)] [Medline: [24973083](https://pubmed.ncbi.nlm.nih.gov/24973083/)]
27. Sangster J, Furber S, Allman-Farinelli M, Phongsavan P, Redfern J, Haas M, et al. Effectiveness of a pedometer-based telephone coaching program on weight and physical activity for people referred to a cardiac rehabilitation program: a randomized controlled trial. *J Cardiopulm Rehabil Prev* 2015;35(2):124-129. [doi: [10.1097/HCR.0000000000000082](https://doi.org/10.1097/HCR.0000000000000082)] [Medline: [25313450](https://pubmed.ncbi.nlm.nih.gov/25313450/)]
28. Axley P, Kodali S, Kuo YF, 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;38(5):924-931. [doi: [10.1111/liv.13622](https://doi.org/10.1111/liv.13622)] [Medline: [29117472](https://pubmed.ncbi.nlm.nih.gov/29117472/)]
29. Aschbrenner KA, Naslund JA, Shevenell M, Mueser KT, Bartels SJ. Feasibility of behavioral weight loss treatment enhanced with peer support and mobile health technology for individuals with serious mental illness. *Psychiatr Q* 2016;87(3):401-415 [FREE Full text] [doi: [10.1007/s11126-015-9395-x](https://doi.org/10.1007/s11126-015-9395-x)] [Medline: [26462674](https://pubmed.ncbi.nlm.nih.gov/26462674/)]

30. Naslund JA, Aschbrenner KA, Marsch LA, McHugo GJ, Bartels SJ. Facebook for supporting a lifestyle intervention for people with major depressive disorder, bipolar disorder, and schizophrenia: an exploratory study. *Psychiatr Q* 2018;89(1):81-94 [FREE Full text] [doi: [10.1007/s11126-017-9512-0](https://doi.org/10.1007/s11126-017-9512-0)] [Medline: [28470468](https://pubmed.ncbi.nlm.nih.gov/28470468/)]
31. Batsis JA, Petersen CL, Clark MM, Cook SB, Kotz D, Gooding TL, et al. Feasibility and acceptability of a technology-based, rural weight management intervention in older adults with obesity. *BMC Geriatr* 2021;21(1):44 [FREE Full text] [doi: [10.1186/s12877-020-01978-x](https://doi.org/10.1186/s12877-020-01978-x)] [Medline: [33435877](https://pubmed.ncbi.nlm.nih.gov/33435877/)]
32. Cotie LM, Prince SA, Elliott CG, Ziss MC, McDonnell LA, Mullen KA, et al. The effectiveness of eHealth interventions on physical activity and measures of obesity among working-age women: a systematic review and meta-analysis. *Obes Rev* 2018;19(10):1340-1358. [doi: [10.1111/obr.12700](https://doi.org/10.1111/obr.12700)] [Medline: [30156044](https://pubmed.ncbi.nlm.nih.gov/30156044/)]
33. Seo DC, Niu J. Evaluation of Internet-based interventions on waist circumference reduction: a meta-analysis. *J Med Internet Res* 2015;17(7):e181 [FREE Full text] [doi: [10.2196/jmir.3921](https://doi.org/10.2196/jmir.3921)] [Medline: [26199208](https://pubmed.ncbi.nlm.nih.gov/26199208/)]
34. Wieland LS, 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;8(8):CD007675 [FREE Full text] [doi: [10.1002/14651858.CD007675.pub2](https://doi.org/10.1002/14651858.CD007675.pub2)] [Medline: [22895964](https://pubmed.ncbi.nlm.nih.gov/22895964/)]
35. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. *Obes Rev* 2015;16(5):376-392. [doi: [10.1111/obr.12268](https://doi.org/10.1111/obr.12268)] [Medline: [25753009](https://pubmed.ncbi.nlm.nih.gov/25753009/)]
36. Joiner KL, Nam S, Whittemore R. Lifestyle interventions based on the diabetes prevention program delivered via eHealth: a systematic review and meta-analysis. *Prev Med* 2017;100:194-207 [FREE Full text] [doi: [10.1016/j.ypmed.2017.04.033](https://doi.org/10.1016/j.ypmed.2017.04.033)] [Medline: [28456513](https://pubmed.ncbi.nlm.nih.gov/28456513/)]
37. Levine DM, Savarimuthu S, Squires A, Nicholson J, Jay M. Technology-assisted weight loss interventions in primary care: a systematic review. *J Gen Intern Med* 2015;30(1):107-117. [doi: [10.1007/s11606-014-2987-6](https://doi.org/10.1007/s11606-014-2987-6)] [Medline: [25134692](https://pubmed.ncbi.nlm.nih.gov/25134692/)]
38. Arem H, Irwin M. A review of web-based weight loss interventions in adults. *Obes Rev* 2011;12(5):e236-e243 [FREE Full text] [doi: [10.1111/j.1467-789X.2010.00787.x](https://doi.org/10.1111/j.1467-789X.2010.00787.x)] [Medline: [20804523](https://pubmed.ncbi.nlm.nih.gov/20804523/)]
39. Anderson-Bill ES, Winett RA, Wojcik JR, Winett SG. Web-based guide to health: relationship of theoretical variables to change in physical activity, nutrition and weight at 16-months. *J Med Internet Res* 2011;13(1):e27 [FREE Full text] [doi: [10.2196/jmir.1614](https://doi.org/10.2196/jmir.1614)] [Medline: [21447470](https://pubmed.ncbi.nlm.nih.gov/21447470/)]
40. Anderson-Bill ES, Winett RA, Wojcik JR. Social cognitive determinants of nutrition and physical activity among web-health users enrolling in an online intervention: the influence of social support, self-efficacy, outcome expectations, and self-regulation. *J Med Internet Res* 2011;13(1):e28 [FREE Full text] [doi: [10.2196/jmir.1551](https://doi.org/10.2196/jmir.1551)] [Medline: [21441100](https://pubmed.ncbi.nlm.nih.gov/21441100/)]
41. Spahn JM, Reeves RS, Keim KS, Laquatra I, Kellogg M, Jortberg B, et al. State of the evidence regarding behavior change theories and strategies in nutrition counseling to facilitate health and food behavior change. *J Am Diet Assoc* 2010;110(6):879-891. [doi: [10.1016/j.jada.2010.03.021](https://doi.org/10.1016/j.jada.2010.03.021)] [Medline: [20497777](https://pubmed.ncbi.nlm.nih.gov/20497777/)]
42. 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;151(4):264-269 [FREE Full text] [doi: [10.7326/0003-4819-151-4-200908180-00135](https://doi.org/10.7326/0003-4819-151-4-200908180-00135)] [Medline: [19622511](https://pubmed.ncbi.nlm.nih.gov/19622511/)]
43. Finlex data bank. URL: <https://www.finlex.fi/en/> [accessed 2021-02-21]
44. 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](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]
45. Furlan AD, Malmivaara A, Chou R, Maher CG, Deyo RA, Schoene M, Editorial Board of the Cochrane Back, Neck Group. 2015 updated method guideline for systematic reviews in the Cochrane back and neck group. *Spine (Phila Pa 1976)* 2015;40(21):1660-1673. [doi: [10.1097/BRS.0000000000001061](https://doi.org/10.1097/BRS.0000000000001061)] [Medline: [26208232](https://pubmed.ncbi.nlm.nih.gov/26208232/)]
46. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005;5:13. [doi: [10.1186/1471-2288-5-13](https://doi.org/10.1186/1471-2288-5-13)] [Medline: [15840177](https://pubmed.ncbi.nlm.nih.gov/15840177/)]
47. R Core Team. R: a language and environment for statistical computing. R foundation for statistical computing. 2019. URL: <https://www.R-project.org/> [accessed 2017-11-06]
48. Higgins J, Green S. Cochrane handbook for systematic reviews interventions. Version 5.1.0. Cochrane training. 2011. URL: <http://www.handbook.cochrane.org> [accessed 2017-11-10]
49. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629-634 [FREE Full text] [doi: [10.1136/bmj.315.7109.629](https://doi.org/10.1136/bmj.315.7109.629)] [Medline: [9310563](https://pubmed.ncbi.nlm.nih.gov/9310563/)]
50. Konstantopoulos S. Fixed effects and variance components estimation in three-level meta-analysis. *Res Synth Methods* 2011;2(1):61-76. [doi: [10.1002/jrsm.35](https://doi.org/10.1002/jrsm.35)] [Medline: [26061600](https://pubmed.ncbi.nlm.nih.gov/26061600/)]
51. Gleser LJ, Olkin I. Stochastically dependent effect sizes. In: Cooper HM, Hedges LV, Valentine JC, editors. *The handbook of research synthesis and meta-analysis*. 2nd edition. New York: Russell Sage Foundation; 2009:357-376.
52. Cai X, Qiu SH, Yin H, Sun ZL, Ju CP, Zügel M, et al. Pedometer intervention and weight loss in overweight and obese adults with type 2 diabetes: a meta-analysis. *Diabet Med* 2016;33(8):1035-1044 [FREE Full text] [doi: [10.1111/dme.13104](https://doi.org/10.1111/dme.13104)] [Medline: [26926674](https://pubmed.ncbi.nlm.nih.gov/26926674/)]

53. Svetkey LP, Clark JM, Funk K, Corsino L, Batch BC, Hollis JF, et al. Greater weight loss with increasing age in the weight loss maintenance trial. *Obesity (Silver Spring)* 2014;22(1):39-44 [FREE Full text] [doi: [10.1002/oby.20506](https://doi.org/10.1002/oby.20506)] [Medline: [23640912](https://pubmed.ncbi.nlm.nih.gov/23640912/)]
54. Morgan PJ, Scott HA, Young MD, Plotnikoff RC, Collins CE, Callister R. Associations between program outcomes and adherence to Social Cognitive theory tasks: process evaluation of the SHED-IT community weight loss trial for men. *Int J Behav Nutr Phys Act* 2014;11:89 [FREE Full text] [doi: [10.1186/s12966-014-0089-9](https://doi.org/10.1186/s12966-014-0089-9)] [Medline: [25011421](https://pubmed.ncbi.nlm.nih.gov/25011421/)]
55. Stephens J, Allen J. Mobile phone interventions to increase physical activity and reduce weight: a systematic review. *J Cardiovasc Nurs* 2013;28(4):320-329 [FREE Full text] [doi: [10.1097/JCN.0b013e318250a3e7](https://doi.org/10.1097/JCN.0b013e318250a3e7)] [Medline: [22635061](https://pubmed.ncbi.nlm.nih.gov/22635061/)]
56. 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(12):CD009386. [doi: [10.1002/14651858.CD009386.pub2](https://doi.org/10.1002/14651858.CD009386.pub2)] [Medline: [26691216](https://pubmed.ncbi.nlm.nih.gov/26691216/)]
57. Pal K, Eastwood SV, Michie S, Farmer AJ, Barnard ML, Peacock R, et al. Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2013;2013(3):CD008776. [doi: [10.1002/14651858.CD008776.pub2](https://doi.org/10.1002/14651858.CD008776.pub2)] [Medline: [23543567](https://pubmed.ncbi.nlm.nih.gov/23543567/)]
58. Kislring LA, Das JM. Prevention strategies. Treasure Island: StatPearls Publishing; 2021.
59. Richardson CR, Newton TL, Abraham JJ, Sen A, Jimbo M, Swartz AM. A meta-analysis of pedometer-based walking interventions and weight loss. *Ann Fam Med* 2008;6(1):69-77 [FREE Full text] [doi: [10.1370/afm.761](https://doi.org/10.1370/afm.761)] [Medline: [18195317](https://pubmed.ncbi.nlm.nih.gov/18195317/)]
60. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. *Am J Prev Med* 2007;33(4):336-345 [FREE Full text] [doi: [10.1016/j.amepre.2007.05.007](https://doi.org/10.1016/j.amepre.2007.05.007)] [Medline: [17888860](https://pubmed.ncbi.nlm.nih.gov/17888860/)]
61. Aguilar-Martínez A, Solé-Sedeño JM, Mancebo-Moreno G, Medina FX, Carreras-Collado R, Saigí-Rubió F. Use of mobile phones as a tool for weight loss: a systematic review. *J Telemed Telecare* 2014;20(6):339-349. [doi: [10.1177/1357633X14537777](https://doi.org/10.1177/1357633X14537777)] [Medline: [24875928](https://pubmed.ncbi.nlm.nih.gov/24875928/)]
62. Johns DJ, Hartmann-Boyce J, Jebb SA, Aveyard P, Behavioural Weight Management Review Group. Weight change among people randomized to minimal intervention control groups in weight loss trials. *Obesity (Silver Spring)* 2016;24(4):772-780 [FREE Full text] [doi: [10.1002/oby.21255](https://doi.org/10.1002/oby.21255)] [Medline: [27028279](https://pubmed.ncbi.nlm.nih.gov/27028279/)]
63. 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;53(6):953-967. [doi: [10.23736/S1973-9087.17.04585-3](https://doi.org/10.23736/S1973-9087.17.04585-3)] [Medline: [28466628](https://pubmed.ncbi.nlm.nih.gov/28466628/)]
64. Aguiar EJ, Morgan PJ, Collins CE, Plotnikoff RC, Young MD, Callister R. Efficacy of the type 2 diabetes prevention using lifestyle education program RCT. *Am J Prev Med* 2016;50(3):353-364. [doi: [10.1016/j.amepre.2015.08.020](https://doi.org/10.1016/j.amepre.2015.08.020)] [Medline: [26526160](https://pubmed.ncbi.nlm.nih.gov/26526160/)]
65. Anderson DR, Christison-Lagay J, Villagra V, Liu H, Dziura J. Managing the space between visits: a randomized trial of disease management for diabetes in a community health center. *J Gen Intern Med* 2010;25(10):1116-1122 [FREE Full text] [doi: [10.1007/s11606-010-1419-5](https://doi.org/10.1007/s11606-010-1419-5)] [Medline: [20556536](https://pubmed.ncbi.nlm.nih.gov/20556536/)]
66. Chambliss HO, Huber RC, Finley CE, McDoniel SO, Kitzman-Ulrich H, Wilkinson WJ. Computerized self-monitoring and technology-assisted feedback for weight loss with and without an enhanced behavioral component. *Patient Educ Couns* 2011;85(3):375-382. [doi: [10.1016/j.pec.2010.12.024](https://doi.org/10.1016/j.pec.2010.12.024)] [Medline: [21295433](https://pubmed.ncbi.nlm.nih.gov/21295433/)]
67. Cho S, Wee JH, Yoo S, Heo E, Ryu B, Kim Y, et al. Effect of lifestyle modification using a smartphone application on obesity with obstructive sleep apnea: a short-term, randomized controlled study. *Clin Exp Otorhinolaryngol* 2018;11(3):192-198 [FREE Full text] [doi: [10.21053/ceo.2017.01284](https://doi.org/10.21053/ceo.2017.01284)] [Medline: [29374961](https://pubmed.ncbi.nlm.nih.gov/29374961/)]
68. Collins CE, Morgan PJ, Jones P, Fletcher K, Martin J, Aguiar EJ, et al. A 12-week commercial web-based weight-loss program for overweight and obese adults: randomized controlled trial comparing basic versus enhanced features. *J Med Internet Res* 2012;14(2):e57 [FREE Full text] [doi: [10.2196/jmir.1980](https://doi.org/10.2196/jmir.1980)] [Medline: [22555246](https://pubmed.ncbi.nlm.nih.gov/22555246/)]
69. Hageman PA, Pullen CH, Hertzog M, Boeckner LS. Effectiveness of tailored lifestyle interventions, using web-based and print-mail, for reducing blood pressure among rural women with prehypertension: main results of the Wellness for Women: DASHing towards Health clinical trial. *Int J Behav Nutr Phys Act* 2014;11:148 [FREE Full text] [doi: [10.1186/s12966-014-0148-2](https://doi.org/10.1186/s12966-014-0148-2)] [Medline: [25480461](https://pubmed.ncbi.nlm.nih.gov/25480461/)]
70. Hansen AW, Grønbaek M, Helge JW, Severin M, Curtis T, Tolstrup JS. Effect of a Web-based intervention to promote physical activity and improve health among physically inactive adults: a population-based randomized controlled trial. *J Med Internet Res* 2012;14(5):e145 [FREE Full text] [doi: [10.2196/jmir.2109](https://doi.org/10.2196/jmir.2109)] [Medline: [23111127](https://pubmed.ncbi.nlm.nih.gov/23111127/)]
71. 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;16:323 [FREE Full text] [doi: [10.1186/s13063-015-0880-1](https://doi.org/10.1186/s13063-015-0880-1)] [Medline: [26223309](https://pubmed.ncbi.nlm.nih.gov/26223309/)]
72. Hunter CM, Peterson AL, Alvarez LM, Poston WC, Brundige AR, Haddock CK, et al. Weight management using the internet a randomized controlled trial. *Am J Prev Med* 2008;34(2):119-126. [doi: [10.1016/j.amepre.2007.09.026](https://doi.org/10.1016/j.amepre.2007.09.026)] [Medline: [18201641](https://pubmed.ncbi.nlm.nih.gov/18201641/)]

73. Lin PH, Wang Y, Levine E, Askew S, Lin S, Chang C, et al. A text messaging-assisted randomized lifestyle weight loss clinical trial among overweight adults in Beijing. *Obesity (Silver Spring)* 2014;22(5):E29-E37. [doi: [10.1002/oby.20686](https://doi.org/10.1002/oby.20686)] [Medline: [24375969](https://pubmed.ncbi.nlm.nih.gov/24375969/)]
74. Mehring M, Haag M, Linde K, Wagenpfeil S, Frensch F, Blome J, et al. Effects of a general practice guided web-based weight reduction program--results of a cluster-randomized controlled trial. *BMC Fam Pract* 2013;14:76 [FREE Full text] [doi: [10.1186/1471-2296-14-76](https://doi.org/10.1186/1471-2296-14-76)] [Medline: [23981507](https://pubmed.ncbi.nlm.nih.gov/23981507/)]
75. Melchart D, Löw P, Wühr E, Kehl V, Weidenhammer W. Effects of a tailored lifestyle self-management intervention (TALENT) study on weight reduction: a randomized controlled trial. *Diabetes Metab Syndr Obes* 2017;10:235-245 [FREE Full text] [doi: [10.2147/DMSO.S135572](https://doi.org/10.2147/DMSO.S135572)] [Medline: [28684917](https://pubmed.ncbi.nlm.nih.gov/28684917/)]
76. Morgan PJ, Callister R, Collins CE, Plotnikoff RC, Young MD, Berry N, et al. The SHED-IT community trial: a randomized controlled trial of internet- and paper-based weight loss programs tailored for overweight and obese men. *Ann Behav Med* 2013;45(2):139-152. [doi: [10.1007/s12160-012-9424-z](https://doi.org/10.1007/s12160-012-9424-z)] [Medline: [23129021](https://pubmed.ncbi.nlm.nih.gov/23129021/)]
77. Rimmer JH, Wang E, Pellegrini CA, Lullo C, Gerber BS. Telehealth weight management intervention for adults with physical disabilities: a randomized controlled trial. *Am J Phys Med Rehabil* 2013;92(12):1084-1094. [doi: [10.1097/PHM.0b013e31829e780e](https://doi.org/10.1097/PHM.0b013e31829e780e)] [Medline: [24257266](https://pubmed.ncbi.nlm.nih.gov/24257266/)]
78. Rogers RJ, Lang W, Barone Gibbs B, Davis KK, Burke LE, Kovacs SJ, et al. Applying a technology-based system for weight loss in adults with obesity. *Obes Sci Pract* 2016;2(1):3-12 [FREE Full text] [doi: [10.1002/osp4.18](https://doi.org/10.1002/osp4.18)] [Medline: [27812375](https://pubmed.ncbi.nlm.nih.gov/27812375/)]
79. Sakane N, Dohi S, Sakata K, Hagiwara SI, Morimoto T, Uchida T, J-Value Study Group. Effects of visceral fat accumulation awareness on a Web-based weight-loss program: Japanese study of visceral adiposity and lifestyle information-utilization and evaluation (J-VALUE). *ISRN Obes* 2013;2013:473764 [FREE Full text] [doi: [10.1155/2013/473764](https://doi.org/10.1155/2013/473764)] [Medline: [24555144](https://pubmed.ncbi.nlm.nih.gov/24555144/)]
80. Shuger SL, Barry VW, Sui X, McClain A, Hand GA, Wilcox S, et al. Electronic feedback in a diet- and physical activity-based lifestyle intervention for weight loss: a randomized controlled trial. *Int J Behav Nutr Phys Act* 2011;8:41 [FREE Full text] [doi: [10.1186/1479-5868-8-41](https://doi.org/10.1186/1479-5868-8-41)] [Medline: [21592351](https://pubmed.ncbi.nlm.nih.gov/21592351/)]
81. Stephens JD, Yager AM, Allen J. Smartphone technology and text messaging for weight loss in young adults: a randomized controlled trial. *J Cardiovasc Nurs* 2017;32(1):39-46. [doi: [10.1097/JCN.0000000000000307](https://doi.org/10.1097/JCN.0000000000000307)] [Medline: [26646593](https://pubmed.ncbi.nlm.nih.gov/26646593/)]
82. Watson S, Woodside JV, Ware LJ, Hunter SJ, McGrath A, Cardwell CR, et al. Effect of a Web-based behavior change program on weight loss and cardiovascular risk factors in overweight and obese adults at high risk of developing cardiovascular disease: randomized controlled trial. *J Med Internet Res* 2015;17(7):e177 [FREE Full text] [doi: [10.2196/jmir.3828](https://doi.org/10.2196/jmir.3828)] [Medline: [26183659](https://pubmed.ncbi.nlm.nih.gov/26183659/)]
83. Chen YC, Tsao LI, Huang CH, Yu YY, Liu IL, Jou HJ. An Internet-based health management platform may effectively reduce the risk factors of metabolic syndrome among career women. *Taiwan J Obstet Gynecol* 2013;52(2):215-221. [doi: [10.1016/j.tjog.2013.04.011](https://doi.org/10.1016/j.tjog.2013.04.011)] [Medline: [23915854](https://pubmed.ncbi.nlm.nih.gov/23915854/)]
84. Eakin EG, Winkler EA, Dunstan DW, Healy GN, Owen N, Marshall AM, et al. Living well with diabetes: 24-month outcomes from a randomized trial of telephone-delivered weight loss and physical activity intervention to improve glycemic control. *Diabetes Care* 2014;37(8):2177-2185. [doi: [10.2337/dc13-2427](https://doi.org/10.2337/dc13-2427)] [Medline: [24658390](https://pubmed.ncbi.nlm.nih.gov/24658390/)]
85. Haapala I, Barengo NC, Biggs S, Surakka L, Manninen P. Weight loss by mobile phone: a 1-year effectiveness study. *Public Health Nutr* 2009;12(12):2382-2391. [doi: [10.1017/S1368980009005230](https://doi.org/10.1017/S1368980009005230)] [Medline: [19323865](https://pubmed.ncbi.nlm.nih.gov/19323865/)]
86. Harrigan M, Cartmel B, Loftfield E, Sanft T, Chagpar AB, Zhou Y, et al. Randomized trial comparing telephone versus in-person weight loss counseling on body composition and circulating biomarkers in women treated for breast cancer: the lifestyle, exercise, and nutrition (LEAN) study. *J Clin Oncol* 2016;34(7):669-676 [FREE Full text] [doi: [10.1200/JCO.2015.61.6375](https://doi.org/10.1200/JCO.2015.61.6375)] [Medline: [26598750](https://pubmed.ncbi.nlm.nih.gov/26598750/)]
87. Karhula T, Vuorinen AL, Rääpysjärvi K, Pakanen M, Itkonen P, Tepponen M, et al. Telemonitoring and mobile phone-based health coaching among Finnish diabetic and heart disease patients: randomized controlled trial. *J Med Internet Res* 2015;17(6):e153 [FREE Full text] [doi: [10.2196/jmir.4059](https://doi.org/10.2196/jmir.4059)] [Medline: [26084979](https://pubmed.ncbi.nlm.nih.gov/26084979/)]
88. Ligibel JA, Meyerhardt J, Pierce JP, Najita J, Shockro L, Campbell N, et al. Impact of a telephone-based physical activity intervention upon exercise behaviors and fitness in cancer survivors enrolled in a cooperative group setting. *Breast Cancer Res Treat* 2012;132(1):205-213 [FREE Full text] [doi: [10.1007/s10549-011-1882-7](https://doi.org/10.1007/s10549-011-1882-7)] [Medline: [22113257](https://pubmed.ncbi.nlm.nih.gov/22113257/)]
89. Reeves M, Winkler E, McCarthy N, Lawler S, Terranova C, Hayes S, et al. The Living Well after Breast Cancer™ pilot trial: a weight loss intervention for women following treatment for breast cancer. *Asia Pac J Clin Oncol* 2017;13(3):125-136. [doi: [10.1111/ajco.12629](https://doi.org/10.1111/ajco.12629)] [Medline: [27863058](https://pubmed.ncbi.nlm.nih.gov/27863058/)]
90. Stuart KL, Wyld B, Bastiaans K, Stocks N, Brinkworth G, Mohr P, et al. A telephone-supported cardiovascular lifestyle programme (CLIP) for lipid reduction and weight loss in general practice patients: a randomised controlled pilot trial. *Public Health Nutr* 2014;17(3):640-647. [doi: [10.1017/S1368980013000220](https://doi.org/10.1017/S1368980013000220)] [Medline: [23452940](https://pubmed.ncbi.nlm.nih.gov/23452940/)]
91. van Wier MF, Ariëns GA, Dekkers JC, Hendriksen IJ, Smid T, van Mechelen W. Phone and e-mail counselling are effective for weight management in an overweight working population: a randomized controlled trial. *BMC Public Health* 2009;9:6 [FREE Full text] [doi: [10.1186/1471-2458-9-6](https://doi.org/10.1186/1471-2458-9-6)] [Medline: [19134171](https://pubmed.ncbi.nlm.nih.gov/19134171/)]

92. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. *J Med Internet Res* 2014;16(9):e186 [FREE Full text] [doi: [10.2196/jmir.3340](https://doi.org/10.2196/jmir.3340)] [Medline: [25217464](https://pubmed.ncbi.nlm.nih.gov/25217464/)]
93. Matthews CE, Wilcox S, Hanby CL, Der Ananian C, Heiney SP, Gebretsadik T, et al. Evaluation of a 12-week home-based walking intervention for breast cancer survivors. *Support Care Cancer* 2007;15(2):203-211. [doi: [10.1007/s00520-006-0122-x](https://doi.org/10.1007/s00520-006-0122-x)] [Medline: [17001492](https://pubmed.ncbi.nlm.nih.gov/17001492/)]
94. Monthly report: population estimates by age (five year groups) and sex. Statistics Bureau of Japan. URL: <http://www.stat.go.jp/english/data/jinsui/tsuki/index.htm> [accessed 2017-11-06]
95. Age and sex composition in the United States: 2010: detailed tables by age and sex. United States Census Bureau. 2010. URL: <https://www.census.gov/data/tables/2010/demo/age-and-sex/2010-age-sex-composition.html> [accessed 2020-10-14]
96. 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;36(5):675-685. [doi: [10.1038/ijo.2011.121](https://doi.org/10.1038/ijo.2011.121)] [Medline: [21694698](https://pubmed.ncbi.nlm.nih.gov/21694698/)]
97. Khaylis A, Yiaslas T, Bergstrom J, Gore-Felton C. A review of efficacious technology-based weight-loss interventions: five key components. *Telemed J E Health* 2010;16(9):931-938 [FREE Full text] [doi: [10.1089/tmj.2010.0065](https://doi.org/10.1089/tmj.2010.0065)] [Medline: [21091286](https://pubmed.ncbi.nlm.nih.gov/21091286/)]
98. Van Rhoon L, Byrne M, Morrissey E, Murphy J, McSharry J. A systematic review of the behaviour change techniques and digital features in technology-driven type 2 diabetes prevention interventions. *Digit Health* 2020;6:2055207620914427 [FREE Full text] [doi: [10.1177/2055207620914427](https://doi.org/10.1177/2055207620914427)] [Medline: [32269830](https://pubmed.ncbi.nlm.nih.gov/32269830/)]
99. Lemstra M, Bird Y, Nwankwo C, Rogers M, Moraros J. Weight loss intervention adherence and factors promoting adherence: a meta-analysis. *Patient Prefer Adherence* 2016;10:1547-1559 [FREE Full text] [doi: [10.2147/PPA.S103649](https://doi.org/10.2147/PPA.S103649)] [Medline: [27574404](https://pubmed.ncbi.nlm.nih.gov/27574404/)]
100. Nuttall FQ. Body mass index: obesity, BMI, and health: a critical review. *Nutr Today* 2015;50(3):117-128 [FREE Full text] [doi: [10.1097/NT.000000000000092](https://doi.org/10.1097/NT.000000000000092)] [Medline: [27340299](https://pubmed.ncbi.nlm.nih.gov/27340299/)]
101. Han TS, Richmond P, Avenell A, Lean ME. Waist circumference reduction and cardiovascular benefits during weight loss in women. *Int J Obes Relat Metab Disord* 1997;21(2):127-134. [doi: [10.1038/sj.ijo.0800377](https://doi.org/10.1038/sj.ijo.0800377)] [Medline: [9043967](https://pubmed.ncbi.nlm.nih.gov/9043967/)]
102. Liao YS, Li HC, Lu HK, Lai CL, Wang YS, Hsieh KC. Comparison of bioelectrical impedance analysis and dual energy X-ray absorptiometry for total and segmental bone mineral content with a three-compartment model. *Int J Environ Res Public Health* 2020;17(7):2595 [FREE Full text] [doi: [10.3390/ijerph17072595](https://doi.org/10.3390/ijerph17072595)] [Medline: [32290133](https://pubmed.ncbi.nlm.nih.gov/32290133/)]
103. CONSORT - EHEALTH checklist (V.1.6.1): information to include when reporting ehealth/mhealth trials (Web - based/Internet - based intervention and decision aids, but also social media, serious games, DVDs, mobile applications, certain telehealth applications). CONSORT EHEALTH 2011 checklist. 2011. URL: <https://asset.jmir.pub/assets/public/CONSORT-EHEALTH-v1-6.pdf> [accessed 2021-09-05]
104. Anttila MR, Kivistö H, Piirainen A, Kokko K, Malinen A, Pekkonen M, et al. Cardiac rehabilitees' technology experiences before remote rehabilitation: qualitative study using a grounded theory approach. *J Med Internet Res* 2019;21(2):e10985 [FREE Full text] [doi: [10.2196/10985](https://doi.org/10.2196/10985)] [Medline: [30730298](https://pubmed.ncbi.nlm.nih.gov/30730298/)]
105. Anttila MR, Söderlund A, Sjögren T. Patients' experiences of the complex trust-building process within digital cardiac rehabilitation. *PLoS One* 2021;16(3):e0247982. [doi: [10.1371/journal.pone.0247982](https://doi.org/10.1371/journal.pone.0247982)] [Medline: [33690681](https://pubmed.ncbi.nlm.nih.gov/33690681/)]

Abbreviations

AICc: corrected Akaike information criterion

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

DXA: dual-energy x-ray absorptiometry

MD: mean difference

PICOS: Population, Intervention, Comparison, Outcomes, and Study Design

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

RCT: randomized controlled trial

Edited by JMIRPE Office; submitted 28.11.20; peer-reviewed by K Joiner, A Furlan; comments to author 11.01.21; revised version received 06.03.21; accepted 28.12.21; published 24.03.22.

Please cite as:

Lahtio H, Rintala A, Immonen J, Sjögren T

The Effectiveness of Physical Activity-Promoting Web- and Mobile-Based Distance Weight Loss Interventions on Body Composition in Rehabilitation Settings: Systematic Review, Meta-analysis, and Meta-Regression Analysis

J Med Internet Res 2022;24(3):e25906

URL: <https://www.jmir.org/2022/3/e25906>

doi: [10.2196/25906](https://doi.org/10.2196/25906)

PMID: [35323126](https://pubmed.ncbi.nlm.nih.gov/35323126/)

©Heli Lahtio, Aki Rintala, Jaakko Immonen, Tuulikki Sjögren. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 24.03.2022. 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.

Review

Role of Telemedicine in Inflammatory Bowel Disease: Systematic Review and Meta-analysis of Randomized Controlled Trials

Lanlan Pang^{1,2}, MD; Hengyu Liu^{2,3}, MD; Zhidong Liu², MD; Jinyu Tan¹, MD; Long-yuan Zhou¹, MD; Yun Qiu¹, MD; Xiaoqing Lin¹, MD; Jinshen He¹, MD; Xuehua Li⁴, MD; Sinan Lin¹, MD; Subrata Ghosh⁵, MD; Ren Mao^{1*}, MD, PhD; Minhu Chen^{1*}, MD, PhD

¹Department of Gastroenterology, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

²Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

³Guangdong Provincial Key Laboratory of Orthopedics and Traumatology, Department of Spine Surgery, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

⁴Department of Radiology, The First Affiliated Hospital of Sun Yat-sen University, Guangzhou, China

⁵NIHR Biomedical Research Centre, Institute of Translational Medicine, University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom

*these authors contributed equally

Corresponding Author:

Ren Mao, MD, PhD

Department of Gastroenterology

The First Affiliated Hospital of Sun Yat-sen University

58 Zhongshan II Road

Guangzhou, 510080

China

Phone: 86 13286834857

Email: maor5@mail.sysu.edu.cn

Abstract

Background: Telemedicine plays an important role in the management of inflammatory bowel disease (IBD), particularly during a pandemic such as COVID-19. However, the effectiveness and efficiency of telemedicine in managing IBD are unclear.

Objective: This systematic review and meta-analysis aimed to compare the impact of telemedicine with that of standard care on the management of IBD.

Methods: We systematically searched the PubMed, Cochrane Library, EMBASE, Web of Science, and Scopus databases on April 22, 2020. Randomized controlled trials comparing telemedicine with standard care in patients with IBD were included, while conference abstracts, letters, reviews, laboratory studies, and case reports were excluded. The IBD-specific quality of life (QoL), disease activity, and remission rate in patients with IBD were assessed as primary outcomes, and the number of in-person clinic visits per patient, patient satisfaction, psychological outcome, and medication adherence were assessed as secondary outcomes. Review Manager 5.3 and Stata 15.1 were used for data analysis.

Results: A total of 17 randomized controlled trials (2571 participants) were included in this meta-analysis. The telemedicine group had higher IBD-specific QoL than the standard care group (standard mean difference 0.18, 95% CI 0.01 to 0.34; $P=0.03$). The number of clinic visits per patient in the telemedicine group was significantly lower than that in the standard care group (standard mean difference -0.71 , 95% CI -1.07 to -0.36 ; $P<.001$). Subgroup analysis showed that adolescents in the telemedicine group had significantly higher IBD-specific QoL than those in the standard care group (standard mean difference 0.42, 95% CI 0.15 to 0.69; $I^2=0$; $P=0.002$), but there was no significant difference between adults in the 2 groups. There were no significant differences in disease activity, remission rate, patient satisfaction, depression, self-efficacy, generic QoL, and medication adherence outcomes between the telemedicine and standard care groups.

Conclusions: Telemedicine intervention showed a promising role in improving IBD-specific QoL among adolescents and decreased the number of clinic visits among patients with IBD. Further research is warranted to identify the group of patients with IBD who would most benefit from telemedicine.

(*J Med Internet Res* 2022;24(3):e28978) doi:[10.2196/28978](https://doi.org/10.2196/28978)

KEYWORDS

telemedicine; inflammatory bowel disease; quality of life; disease activity; mobile phone

Introduction

Inflammatory bowel disease (IBD), including Crohn disease and ulcerative colitis, is a group of chronic inflammatory disorders of the gut. The prevalence of IBD is increasing worldwide, with 3 million cases recorded in the United States in 2015 and 4 million cases projected in Canada by 2030 [1,2]. Because of its recurrent relapsing-remitting nature, IBD exerts a substantial economic and health burden on patients and their families, health organizations, and nations [3,4]. The lack of curative therapy for this condition entails lifelong medication and follow-up that need to be effectively monitored in patients with IBD [5].

Telemedicine was first defined by the World Health Organization as health care service provided to patients at a distance through information communication technologies (ie, SMS text messaging, web-based applications, real-time telephone) [6]. It is a broad term. Although the specific telemedicine subtypes (telemonitoring, tele-education, and teleconsulting) exhibit significant heterogeneity, they are closely tied together by the concept of remote health care resources delivery [7]. Given the convenience of communication technologies, clinicians have been increasingly using eHealth interventions as a supplementary tool to conduct follow-up and provide education, including disease status and medication instruction. Electronic medical technology has been proven to change the course of certain chronic diseases such as diabetes and asthma [8-11]. Patients with IBD, commonly diagnosed as having the condition at a young age and deemed to need lifelong follow-up for long-term remission, could also potentially benefit from telemedicine intervention for preventing disease progression and reducing complications and operation rates [12-14]. Telemedicine has played an important role in the management of IBD during the recent COVID-19 pandemic [15]. Specific tools such as the IBD Monitoring Index for Mobile Health have been developed and validated for clinical management [16-21]. Others tools such as the IBD disk have been adapted to smartphone apps to monitor IBD-associated disability [22,23].

However, there is no consensus on remote health care technology preferences for IBD management because of contradictory results and high heterogeneity among studies. Few studies precisely quantified the magnitude of intervention effects [24-26], although many studies demonstrated that telemedicine had a major impact on the management of IBD

[7,27,28]. We aimed to estimate the differences between telemedicine and standard care in the management of IBD by conducting a systematic review and meta-analysis of randomized controlled trials (RCTs).

Methods

This study was performed in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.

Search Strategy and Selection Criteria**Literature Search Strategy**

Two investigators (LLP and ZDL) independently searched publications in the PubMed, Cochrane Library, EMBASE, Web of Science, and Scopus databases (search date April 22, 2020) using the following search terms: (telemedicine OR telemonitor OR e-health OR telehealth OR telecommunication OR telemanagement OR telecare OR (telephone monitoring) OR telenursing OR ((remote and short) message service) OR (mobile health) OR (mobile applications) OR teleconsultation) AND ((inflammatory bowel disease) OR (ulcerative colitis) OR (Crohn's disease)). The search in the Web of Science, Cochrane Library, and Scopus Google Scholar databases was limited to titles and abstracts. However, no limitations were applied to the search of PubMed and EMBASE. We also manually searched the reference lists and related literature to identify additional publications. The data sets used in this study can be obtained from the corresponding author on request. Records were imported into EndNote X 9.0 software (Clarivate) to eliminate duplications.

Eligibility Criteria and Study Selection

Two authors (LLP and HYL) independently screened the titles, abstracts, and keywords of the identified articles and selected suitable papers for full review. Disagreements were resolved by a third investigator (ZDL) or by consensus.

The studies included had to meet the following PICOS (participants, interventions, control, outcomes, study design) criteria described in [Textbox 1](#) [6].

The exclusion criteria were as follows: conference abstracts, letters, reviews, laboratory studies, and case reports in which the necessary information could not be extracted; non-English publications; and studies that did not report the outcomes required.

Textbox 1. Inclusion criteria.1. *P (participants)*

Patients diagnosed as having IBD

2. *I (interventions)*

Telemedicine defined as “the use of electronic information and communication technologies for the delivery of health care when there exist distances between patients and health care providers” such as internet, mobile phone applications, and SMS text messaging

3. *C (control)*

Standard care or usual care provided by the medical center according to IBD treatment guidelines

4. *O (outcomes)*

At least one of the following outcomes: inflammatory bowel disease–specific quality of life, disease activity, remission rate, generic quality of life, self-efficacy, depression, medication adherence, patient satisfaction, and the number of clinic visits per patient

5. *S (study design)*

Only randomized controlled trials

Risk of Bias

Two reviewers (LLP and HYL) independently assessed the quality and risk of bias of the included studies using the Cochrane Handbook of Systematic Reviews of Interventions [29]. In addition, the revised Jadad scale was also applied to assess the quality of the included articles [30]. Any disagreement was resolved by the third reviewer (ZDL).

Data Extraction

Two authors (LLP and HYL) independently extracted data, and disagreements were resolved by a third investigator (ZDL). Extracted data included first author, publication year, country, participant characteristics (age, gender, disease type, and disease activity status), intervention, follow-up time, and outcomes. The investigators contacted authors to obtain original data not reported in the published papers. If the number of telemedicine intervention groups was more than one, amalgamation of these groups was performed. If outcomes were reported more than once, the updated data would be evaluated on priority.

Outcomes and Definitions

Primary outcomes in our study included IBD-specific QoL, disease activity, and remission rate. Secondary outcomes included generic QoL, self-efficacy, depression, medication adherence, patient satisfaction, and the number of in-person clinic visits per patient.

Except the number of clinic visits, reported outcomes were measured by specific questionnaires or scales. For instance, IBD-specific QoL was assessed by the IBD questionnaire (IBDQ) [31]. Disease activity was assessed by the Mayo score, Walmsley index, or Seo index for ulcerative colitis or indeterminate colitis and by the Harvey Bradshaw Index for Crohn disease [32–35]. Additionally, remission rate was defined as the proportion of patients in clinical remission at the endpoint or during the intervention [36,37]. Patient satisfaction was defined by scales (eg, consultation satisfaction questionnaire) evaluating the acceptance of care provided by clinical staff [38].

Data Synthesis and Statistical Analysis

All data were analyzed using Review Manager 5.3 (The Cochrane Collaboration) and Stata 15.1 (StataCorp). We used

standardized mean difference (SMD) with 95% CI to calculate continuous variables and relative risk with 95% CI to calculate discontinuous variables. Owing to the heterogeneity between the included studies, we used a random-effects model to assess a relatively more conservative estimate of the 95% CI. Heterogeneity was evaluated using the Cochrane Q statistic and I^2 statistic. Subgroup analysis was conducted if needed, focusing on predefined stratification including the follow-up time (<12 months or not) and patient characteristics (adults or adolescents or patients aged above 18 years or not). Funnel plots, Egger test, and Begg test were used to examine potential publication bias.

Results**Search Process, Study Characteristics, and Quality Assessment**

A total of 1422 articles were identified after searching the databases, and 2 additional articles were included through search of the references. Of the 711 unique studies obtained after removing duplications, 617 irrelevant articles were eliminated and 94 were assessed in full text. The following publications were excluded: 12 articles because of a lack of accessible full text, 13 because of incomplete data, 7 for being unrelated to the topic, and 11 for failure of randomization; 21 conference abstracts; and 13 reviews. Finally, 17 RCTs were considered potentially eligible and comprised 2571 patients from 2010 to 2020; most of these RCTs were conducted in the United States [39–45], followed by the Netherlands [46–48], New Zealand [49,50], Denmark [12,51], the United Kingdom [52], Spain [53], Ireland [12], and Turkey [13]. The process of selecting enrolled studies is shown in Figure 1.

Table 1 summarizes the key characteristics of the included studies and participants. The results of the revised Jadad scale for the enrolled studies are also shown in Table 1 and indicate that 13 identified studies were of high quality (ranging from 5 to 7). Three of the included studies are from the same clinical trial but report different outcomes. The methodological quality of enrolled studies is shown in Figure 2 and Figure 3.

Figure 1. Flow diagram of the selection of enrolled studies.

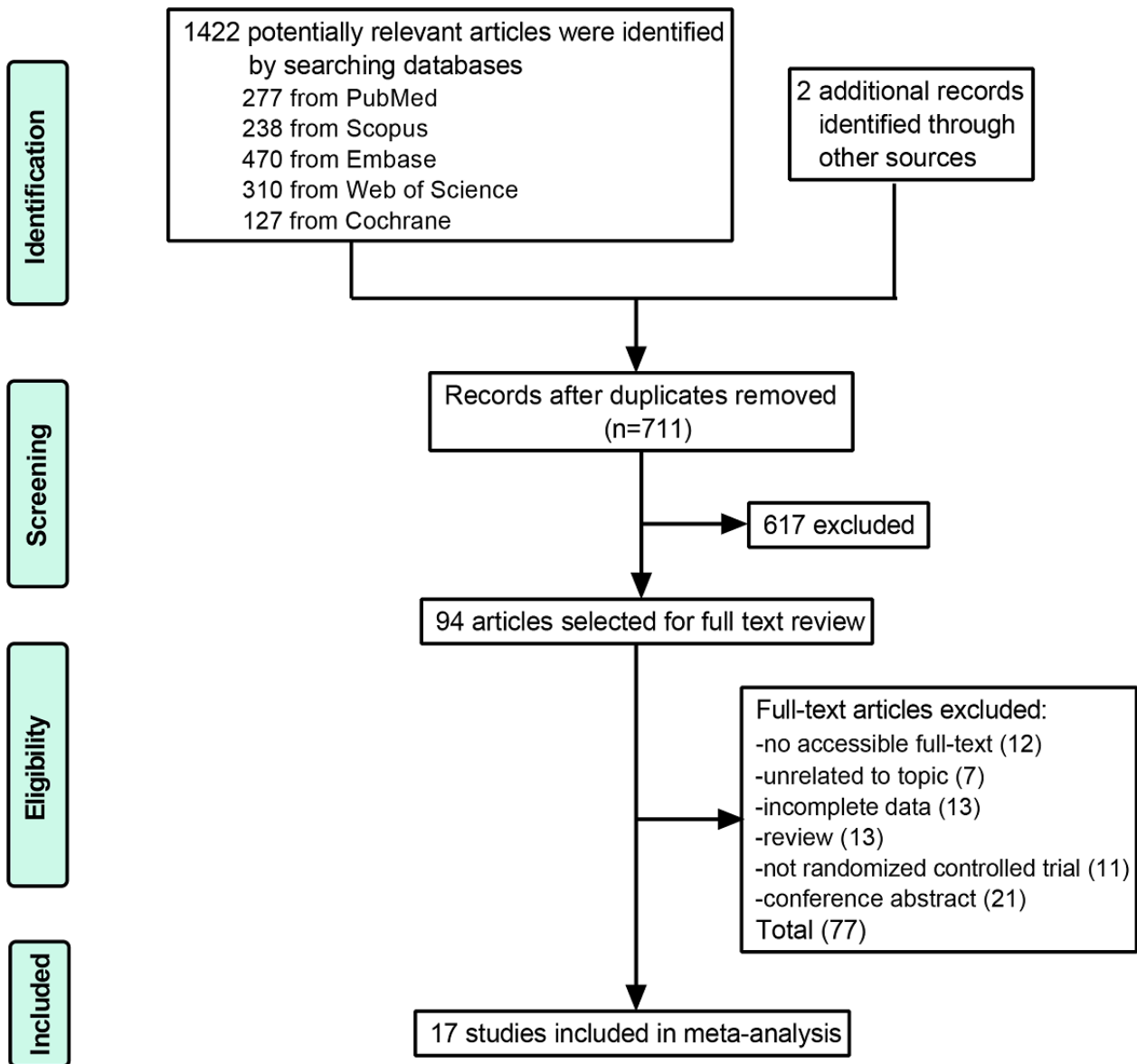


Table 1. Characteristics of the included studies and participants.

Reference	Country	Jadad score	Age (years), mean (SD)/mean (IQR)/mean (range) ^a	Male/total ^a (%)	Participants ^a , n	Participant characteristics	Intervention/application ^a	Follow-up time (months)	Outcomes
Linn et al, 2018 [48]	Netherlands	1+1+0+1=3	40.84 (14.51) vs 45.21(17.15)	40.4 vs 51.5	52 vs 33	<ul style="list-style-type: none"> Individuals diagnosed with IBD^b Individuals receiving immunosuppressant or biological therapy for the first time 	Telemonitoring through web or SMS text messaging vs usual care	6	<ul style="list-style-type: none"> Patient satisfaction Self-efficacy Medication adherence score
Ako Beng et al, 2015 [52]	United Kingdom	2+2+0+1=5	13.9 (12.1,15.9) vs 13.8 (11.2,15.3)	68 vs 57	44 vs 42	<ul style="list-style-type: none"> Young people (aged 8-16 years) with IBD 	Teleconsulting through telephone vs usual care	12	<ul style="list-style-type: none"> IBD-specific QoL^c Patient satisfaction
Bil Grami et al, 2019 ^d [42]	United States	2+2+0+1=5	39.7 (13) vs 37.7 (11.6) vs 40.2 (11)	38.5 vs 47.4 vs 36.9	75 vs 72 vs 75	<ul style="list-style-type: none"> Adults with IBD who experienced an IBD flare within 2 years prior to the trial Individuals at least 18 years of age 	Telemonitoring and tele-education through mobile phone with SMS text messaging vs standard care	12	<ul style="list-style-type: none"> Self-efficacy
Carlsen, 2017 [51]	Denmark	1+2+1+1=5	15.1 (1.82) vs 14.7 (2.11)	37 vs 46	27 vs 26	<ul style="list-style-type: none"> Children and adolescents, 10-17 years old, diagnosed as having IBD 	Telemonitoring through web-based applications, SMS text messaging, and phone call vs standard care	24	<ul style="list-style-type: none"> Number of clinic visits per patient
Cross et al, 2012 [39]	United States	2+2+0+1=5	41.7 (13.9) vs 40.3 (14.4)	40 vs 32	25 vs 22	<ul style="list-style-type: none"> Adults with ulcerative colitis 	Telemonitoring through home unit-server PC provider vs standard care	12	<ul style="list-style-type: none"> IBD-specific QoL Disease activity Medication adherence rate
Cross, 2018 ^d [40]	United States	2+2+0+1=5	40.1 (13.2) vs 36.4 (11.5) vs 40.1 (11.7)	41.7 vs 43.1 vs 45.3	115 vs 116 vs 117	<ul style="list-style-type: none"> Adults ≥18 years of age diagnosed as having IBD who experienced at least one IBD flare in the 2 years prior to the baseline visit 	Telemonitoring and tele-education through mobile phone with SMS text messaging vs standard care	12	<ul style="list-style-type: none"> IBD-specific QoL Disease activity Remission rate
De Jong, 2017[47]	Netherlands	2+2+0+1=5	44.0 (±14.1) vs 44.1 (14.2)	42 vs 41	465 vs 444	<ul style="list-style-type: none"> Outpatients aged 18-75 years with IBD and without an ileoanal or ileorectal pouch anastomosis 	Telemonitoring through web-based applications on a tablet or smartphone vs standard care	12	<ul style="list-style-type: none"> IBD-specific QoL Number of outpatient visits per patient Medication adherence rate Self-efficacy

Reference	Country	Jadad score	Age (years), mean (SD)/mean (IQR)/mean (range) ^a	Male/total ^a (%)	Participants ^a , n	Participant characteristics	Intervention/application ^a	Follow-up time (months)	Outcomes
Del Hoyo et al, 2018 and 2019 [18,53]	Spain	2+2+0+1=5	41.32(19-66) vs 40.91(24-60) vs 39.31(22-61)	42.9 vs 57.1 vs 57.1	21 vs 21 vs 21	<ul style="list-style-type: none"> Adults ≥18 years of age diagnosed as having IBD for at least 6 months Patients who had complex IBD when immunosuppressants or biologic agents were initiated 	Telemonitoring through a web-based system with smartphone apps or a tablet or through the telephone vs standard care	6	<ul style="list-style-type: none"> IBD-specific QoL Remission rate Disease activity Generic QoL Medication adherence score and rate Patient satisfaction
Elkjaer et al, 2010 [12]	Denmark and Ireland	2+2+2+1=7	Denmark: 40 (21-69) vs 44 (21-69) Ireland: 41 (18-66) vs 46 (19-65)	Denmark: 49.5 vs 31.1 Ireland: 60.8 vs 41.5	Denmark: 105 vs 106 Ireland: 51 vs 41	<ul style="list-style-type: none"> Patients aged 18-69 years who met the international diagnostic criteria for mild to moderate ulcerative colitis and were treated with 5-aminosalicylic acid 	Tele-education through web-based applications vs usual care	12	<ul style="list-style-type: none"> Medication adherence rate Remission rate Number of clinic visits
Heida et al, 2017 [46]	Netherlands	2+2+0+1=5	15 (12-16) vs 15 (13-17)	64 vs 45	84 vs 86	<ul style="list-style-type: none"> Patients aged 10-19 years with IBD in clinical remission at baseline Patients diagnosed as having IBD more than 6 months before enrolment 	Telemonitoring through web-based applications, email, and phone calls vs usual care	13	<ul style="list-style-type: none"> IBD-specific QoL Remission rate
Hunt et al, 2017 [44]	United States	0+0+0+1=1	36 (10) (total participants)	20.6 (total participants)	32 vs 31	<ul style="list-style-type: none"> Patients at least 18 years old who self-reported a previous diagnosis of IBD, according to a medical professional's feedback for IBD patients Patients with secondary irritable bowel syndrome or with a known psychological risk factor for poor health-related QoL in chronic gastrointestinal tract disorders 	Tele-education through cognitive behavioral therapy delivered online vs usual care	1.5	<ul style="list-style-type: none"> IBD-specific QoL Disease activity Depression
Krier et al, 2011 [45]	United States	1+2+1+1=5	62.8 (11.5) vs 58.5 (9.6)	87 vs 68	15 vs 19	<ul style="list-style-type: none"> Patients with IBD who underwent 57 encounters in 9 months 	Teleconsulting through real-time image vs standard care	9	<ul style="list-style-type: none"> Patient satisfaction
		2+2+0+1=5		52 vs 46	50 vs 50			12	

Reference	Country	Jadad score	Age (years), mean (SD)/mean (IQR)/mean (range) ^a	Male/total ^a (%)	Participants ^a , n	Participant characteristics	Intervention/application ^a	Follow-up time (months)	Outcomes
McCombie et al, 2020 [49]	New Zealand		35.2 (12.4) vs 34.3 (12.9)			<ul style="list-style-type: none"> Patients who were 16 years or older with confirmed IBD and who had at least 2 outpatient appointments and <3 disease flares in the past 12 months 	Telemonitoring through smart-phone apps vs standard care		<ul style="list-style-type: none"> IBD-specific QoL
McCombie et al, 2016 [50]	New Zealand	2+0+0+1=3	38.3 (12.8) vs 39.6 (11.8)	33.6 vs 38.4	113 vs 86	<ul style="list-style-type: none"> All adults with IBD aged 18 to 65 years 	Tele-education through computerized cognitive behavioral therapy vs usual care	6	<ul style="list-style-type: none"> IBD-specific QoL Generic QoL Depression
Miloh et al, 2017 [43]	United States	1+1+1+0=3	—	—	21 vs 30	<ul style="list-style-type: none"> Children with IBD who were 8 years and older 	Telemonitoring through SMS text messaging vs standard care	12	<ul style="list-style-type: none"> Medication adherence rate Disease activity Number of clinic visits per patient
Ozgur Soy et al, 2019 [13]	Turkey	2+2+2+1=7	37.26(12.99) vs 41.63(11.85)	56.7 vs 60	30 vs 30	<ul style="list-style-type: none"> Adults aged 18 years or over who were diagnosed as having IBD for 6 months 	Tele-education through web-based applications on the computer or phone vs standard care	2	<ul style="list-style-type: none"> IBD-specific QoL Remission rate
Schliep et al, 2020 ^d [41]	United States	2+2+0+1=5	37.3 (11.6) vs 39.3 (13.4) vs 39.5 (12.0)	45 vs 40.5 vs 37.5	71 vs 74 vs 72	<ul style="list-style-type: none"> Adults who were ≥18 years of age, were diagnosed as having IBD, and experienced at least one IBD flare in the 2 years prior to the baseline visit (an increase in IBD symptoms sufficient to warrant a change in medication dose or addition of a medication) 	Telemonitoring and tele-education through a mobile phone with SMS text messaging vs standard care	12	<ul style="list-style-type: none"> Depressive symptoms Generic QoL

^aThese items were recorded as experimental vs control group.

^bIBD: inflammatory bowel disease.

^cQoL: quality of life.

^dThese studies came from the same clinical trial but reported different outcomes.

Figure 2. Risk of bias graph.

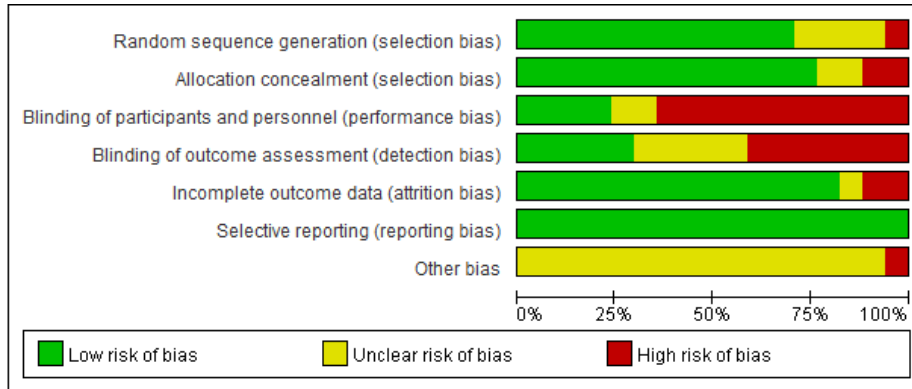


Figure 3. Risk of bias summary.

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
A.J Linn 2018	?	?	+	+	+	+	?
Akobeng 2015	+	+	+	+	+	+	?
Bilgrami 2019	+	+	+	+	+	+	?
Carlsen 2017	?	+	+	?	+	+	?
Cross 2012	+	+	+	+	+	+	?
Cross 2018	+	+	+	+	+	+	?
De Jong 2017	+	+	+	+	+	+	?
Del Hoyo 2018	+	+	+	+	+	+	?
Elkjaer 2010	+	+	+	?	+	+	?
Heida 2017	+	+	+	+	+	+	?
Hunt 2017	+	+	+	+	+	+	?
Krier 2011	?	+	?	?	+	+	?
Mc Combie 2016	+	+	+	+	+	+	?
Mc Combie 2019	+	+	+	+	+	+	?
Miloh 2017	?	?	?	?	?	+	+
Ozgursoy Uran 2019	+	+	+	?	+	+	?
Schliep 2020	+	+	+	+	+	+	?

Primary Outcomes

IBD-Specific Quality of Life

A total of 10 clinical trials including 1632 participants were enrolled to compare IBD-specific QoL in the telemedicine and standard care groups. We found that IBDQ scores were higher in the telemedicine group than in the standard care group (SMD 0.18, 95% CI 0.01 to 0.34; $I^2=47$; $P=.03$; **Figure 4**). Subgroup analysis stratified by follow-up time (<12 months or not) and participants characteristics (adults or adolescents) was conducted to examine the relatively high heterogeneity and identify the type of patients in need of telemedicine care. There was no significant difference in the IBDQ scores in the short-term (SMD

0.23, 95% CI -0.22 to 0.68; $I^2=61$; $P=0.31$) or long-term subgroups (SMD 0.17, 95% CI 0 to 0.34; $I^2=47$; $P=.05$; **Multimedia Appendix 1**). Furthermore, adolescents in the telemedicine group had significantly higher IBDQ scores than those in the standard care group (SMD 0.42, 95% CI 0.15 to 0.69; $I^2=0$; $P=.002$), but no significant difference was found for adults between the groups (SMD 0.11, 95% CI -0.06 to 0.28; $I^2=41$; $P=.21$; **Multimedia Appendix 2**).

Funnel plot showed potential publication bias in our meta-analysis (**Figure 5**), contrary to the results of the Begg ($P=.86$) and Egger test ($P=.26$). This inconformity could be explained by the relatively small number of enrolled studies.

Figure 4. IBD-specific quality of life compared between telemedicine and standard care groups. IBD: inflammatory bowel disease [13,39,40,44,46,47,49,50,52,53].

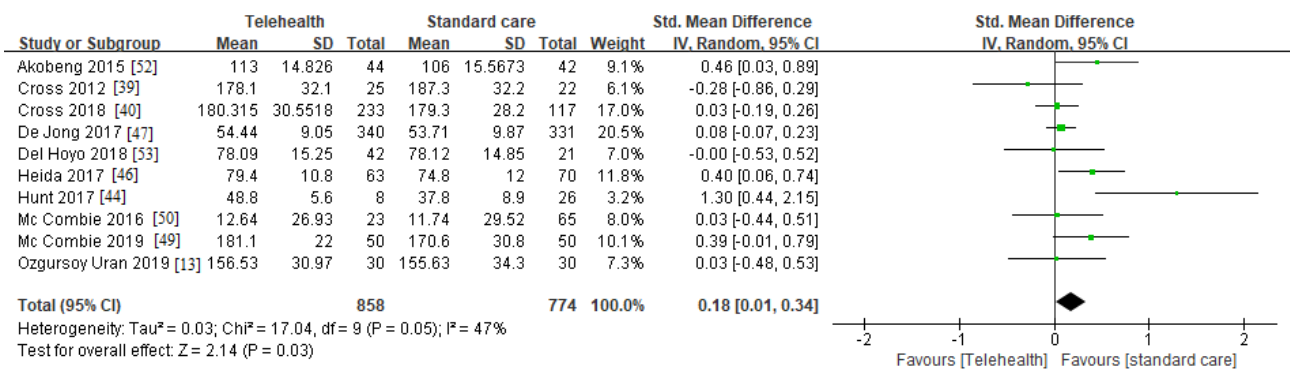
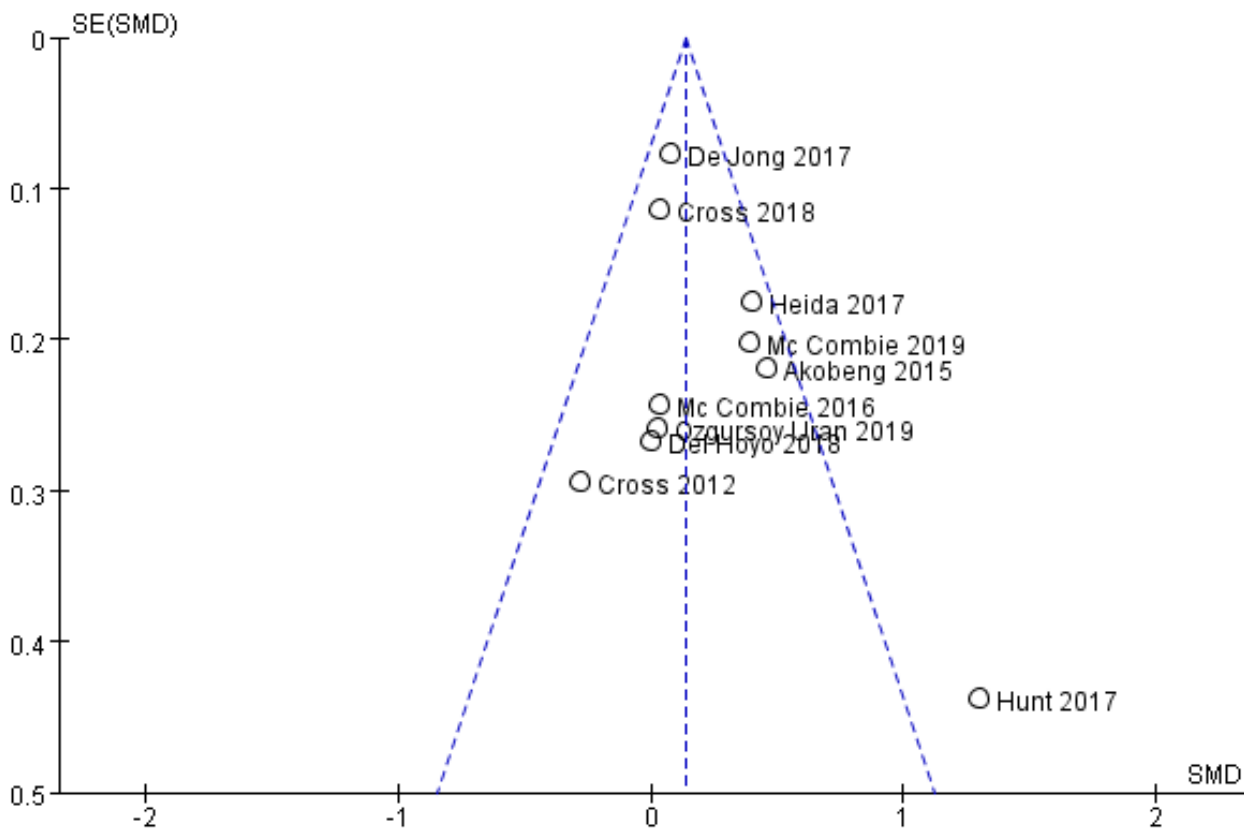


Figure 5. Funnel plot for potential publication bias. SMD: standard mean difference.

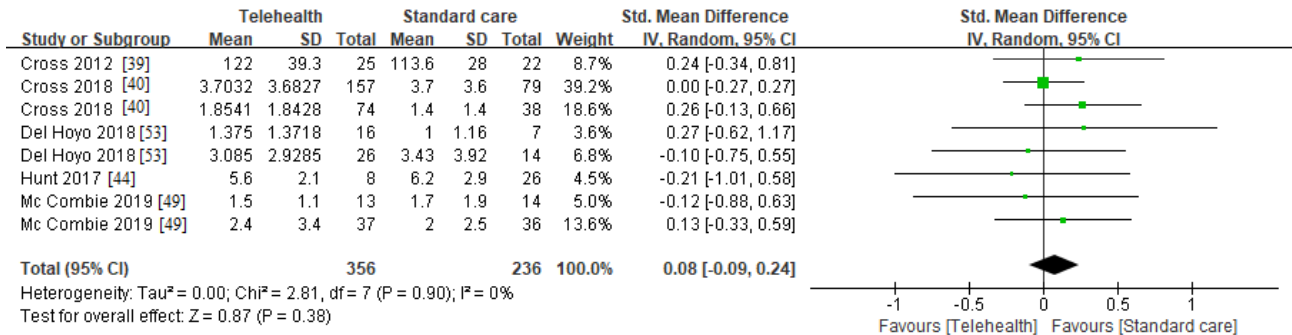


Disease Activity and Remission Rate

To examine the effectiveness of telemedicine in managing disease activity, 7 RCTs with a total of 955 patients were included. Disease activity was not significantly different between the telemedicine and standard care groups (SMD 0.08,

95% CI -0.09 to 0.24; $I^2=0$; $P=.38$; Figure 6). Meanwhile, the remission rate in the telemedicine group was not significantly lower than that in the standard care group (relative risk 0.94, 95% CI, 0.83 to 1.05; $I^2=6$; $P=.26$; Multimedia Appendix 3).

Figure 6. Disease activity in the telemedicine and standard care groups [39,40,44,53].



Secondary Outcomes

Number of Clinic Visits per Patient

To investigate whether telemedicine intervention could lower the number of clinic or outpatient visits, we analyzed 6 articles that included 1479 patients with IBD. The number of clinic visits per patient was significantly lower in the telemedicine group than in the standard care group (SMD -0.71, 95% CI -1.07 to -0.36; $I^2=85$; $P<.001$; Multimedia Appendix 4).

Patient Satisfaction

In 3 studies that included 183 participants, patient satisfaction was not significantly different between the telemedicine and standard care groups (SMD 0.21, 95% CI -0.12 to 0.54; $I^2=14$; $P=0.21$; Multimedia Appendix 5).

Psychological Outcomes (Depression, General QoL, and Self-efficacy)

In the assessment of psychological outcomes, 7 clinical trials with 1165 participants showed no significant difference in the mental health of patients with IBD between the telemedicine and standard care groups (SMD -0.31, 95% CI -0.79 to 0.17; $I^2=66$; $P=.21$ for depression score; SMD 1.37, 95% CI -0.42 to 3.15; $I^2=97$; $P=.13$ for generic QoL; SMD 0.01, 95% CI -0.16 to 0.17, $I^2=23$ %; $P=.95$ for self-efficacy; Multimedia Appendices 6-8).

Medication Compliance

A total of 5 RCTs with 1169 patients with IBD were included to assess medication compliance. Considering that some articles reported the Morisky Scale score while others merely reported the medication compliance rate, we pooled data into 2 measures (medication compliance score and rate). Medication compliance in the telemedicine group did not improve significantly compared with that in the standard care group in terms of medication compliance score and rate (SMD 0.11, 95% CI -0.09 to 0.30; $I^2=19$; $P=.27$ and relative risk 1.29, 95% CI 0.77 to 2.17; $I^2=88$; $P=.33$, respectively; Multimedia Appendix 9).

Discussion

There is high demand for long-term personalized care and medication to maintain remission and reduce the risk of relapse in patients with IBD [5,54]. Because of the convenience of use, telemedicine intervention may play an increasingly important role in managing IBD [27,55]. We aimed to investigate whether patients with IBD could benefit from telemedicine technology by performing a systematic review and meta-analysis. It was evident that enrolled studies exhibited obvious heterogeneity in the specific intervention used. The reason for this heterogeneity could not be identified because of the diversity and physical limitations of the IBD centers delivering telemedicine and their purposes and areas of application. However, regardless of the heterogeneity, we did find that patients who received telemedicine intervention had higher IBDQ scores and a significantly lower number of clinic visits per patient than those who received standard care. Importantly, adolescent patients with IBD benefit more from telemedicine and had significantly higher IBDQ scores than those who received standard care.

One possible reason is that there are more opportunities for the youth to access this relatively new form of care via the internet or mobile phones. Unlike in other chronic diseases (eg, chronic obstructive pulmonary disease), the peak onset of IBD is seen at a younger age [3,56]. This implies that telemedicine would be more acceptable in such patients with IBD. In addition, it seemed obvious that telemedicine could decrease the number of in-person clinic visits compared with standard care. However, none of the studies reported exact data or definitive conclusions on this issue. Considerable time and cost could be saved through the reduction of travel and waiting hours for regular office visits.

Given the robust effects of relapse or disease course on the daily life of patients with IBD, attention should be focused on relieving the psychological burden on these patients [57]. It is necessary to note that telemedicine aiming to improve outcomes in patients with IBD, such as through the incorporation of impactful web-based cognitive behavioral therapy (a form of tele-education), has proven to be an effective method for the

management of mental health in patients with chronic gastrointestinal tract diseases [25]. However, our study showed no significant differences in psychological outcomes, such as depression, generic QoL, and self-efficacy, between the telemedicine and standard care groups. One potential explanation is that standard care provided by the referral center had built in emphasis on the importance of mental health care.

Theoretically, patients receiving telemedicine intervention have more access to report a flare than conventional follow-up and therefore receive more prompt consultation from health care givers. However, no significant differences were observed for disease activity and remission rates between these 2 groups in our meta-analysis. The reasons for this may be as follows: most patients were in remission at baseline, which led to a ceiling effect; it remains uncertain whether eHealth technologies could better influence the natural course of IBD compared with standard care; and evaluation of disease activity was based on the score self-reported by patients or their families without objective measurements. Hence, it is difficult to conclusively state the impact of telemedicine intervention on disease activity in patients with IBD.

In terms of therapeutic compliance, medication adherence was adequate in only around 45% of patients with IBD [58]. Nonadherence to medical therapy could cause a 5-fold increase in the risk of relapse, and low medication compliance correlates with lower QoL and higher cost of hospitalizations [59,60]. Thus, there is an urgent need to promote better medication compliance in patients with IBD. Our findings did not show a significant improvement in medication compliance in the telemedicine group, which was inconsistent with the outcomes reported by Rohde et al [26]. It is reasonable to speculate that this may be related to the compliance rate at baseline, as noncompliant patients might be more reluctant to participate in RCTs. Consequently, the participants enrolled are more compliant with the medication, which results in a ceiling effect.

Because of the superiority and popularity of mobile technology, intervention restricted to mobile phones is considered to be promising for the management of chronic diseases [61,62]. Our meta-analysis not only investigated the effectiveness of telemedicine in IBD, but also focused on the specific type of telemedicine, including mobile technology. All enrolled studies in our meta-analysis incorporated mobile devices into the telemedicine intervention, except 2 in which the intervention was confined to a computerized web-based system [45,50]. Therefore, we anticipate that our findings of the use of telemedicine for the management of IBD could also be applied to mobile technology.

Despite the strengths of this meta-analysis, there are certain limitations. First, there was an unavoidable high attrition rate in some studies that used the per protocol analysis. Second, some RCTs did not use the blinded-method design because of intervention characteristic limitations, which led to performance and detection biases. Third, the number of enrolled studies in the meta-analysis was relatively modest, which led to the contradictory results ascribed to potential publication bias. Finally, the specific population that would most benefit from telemedicine could not be identified because of a lack of complete reported data in the included studies.

In conclusion, constrained by the current limited material to provide telemedicine for IBD patients, the heterogeneity of specific telemedicine intervention was obviously evident. However, in accordance with the idea of providing health care from a distance, telemedicine should not be regarded as a uniform therapeutic method as is done for drug treatments but as a mode of health care delivery and even as an important adjuvant to routine clinical practice. Meanwhile, telemedicine intervention did show a promising role in improving IBDQ scores among adolescents and decreased the number of clinic visits by patients with IBD. Further research is needed to identify the patients with IBD who would most benefit from telemedicine.

Acknowledgments

We would like to thank the National Natural Science Foundation of China for providing funding for this meta-analysis. The funder of this study had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Authors' Contributions

LLP designed the search strategy. LLP and ZDL performed the search. LLP and HYL performed the abstract screening, full-text screening, data extraction, and risk of bias assessment. All the authors drafted and revised the manuscript. RM and MHC supervised the overall study.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Subgroup analysis of the inflammatory bowel disease questionnaire comparing the telemedicine with standard care by follow-up time.

[PNG File, 28 KB - [jmir_v24i3e28978_app1.png](#)]

Multimedia Appendix 2

Subgroup analysis of the inflammatory bowel disease questionnaire comparing the telemedicine with standard care by the age of patients.

[PNG File , 27 KB - [jmir_v24i3e28978_app2.png](#)]

Multimedia Appendix 3

Remission rate comparing the telemedicine versus the standard care.

[PNG File , 12 KB - [jmir_v24i3e28978_app3.png](#)]

Multimedia Appendix 4

The number of clinic visits per patient comparing telemedicine with standard care.

[PNG File , 14 KB - [jmir_v24i3e28978_app4.png](#)]

Multimedia Appendix 5

Patient satisfaction comparing telemedicine with standard care.

[PNG File , 11 KB - [jmir_v24i3e28978_app5.png](#)]

Multimedia Appendix 6

Depression comparing the telemedicine with the standard care.

[PNG File , 11 KB - [jmir_v24i3e28978_app6.png](#)]

Multimedia Appendix 7

Generic quality of life comparing the telemedicine with the standard care.

[PNG File , 11 KB - [jmir_v24i3e28978_app7.png](#)]

Multimedia Appendix 8

Self-efficacy comparing the telemedicine with the standard care.

[PNG File , 12 KB - [jmir_v24i3e28978_app8.png](#)]

Multimedia Appendix 9

(A) Medication compliance score comparing the telemedicine with the standard care.(B) Medication compliance rate comparing the telemedicine with the standard care.

[PNG File , 23 KB - [jmir_v24i3e28978_app9.png](#)]

References

1. Dahlhamer JM, Zammitti EP, Ward BW, Wheaton AG, Croft JB. Prevalence of inflammatory bowel disease among adults aged ≥ 18 years - United States, 2015. MMWR Morb Mortal Wkly Rep 2016 Oct 28;65(42):1166-1169 [FREE Full text] [doi: [10.15585/mmwr.mm6542a3](#)] [Medline: [27787492](#)]
2. Coward S, Clement F, Benchimol E, Bernstein C, Bitton A, Carroll M, et al. Tu1698 - Analyzing the rising prevalence of IBD: predicting the prevalence in 2030 by age group. Gastroenterology 2018 May;154(6):S-992-S-993. [doi: [10.1016/S0016-5085\(18\)33324-9](#)]
3. Ananthakrishnan AN. Epidemiology and risk factors for IBD. Nat Rev Gastroenterol Hepatol 2015 Apr 3;12(4):205-217. [doi: [10.1038/nrgastro.2015.34](#)] [Medline: [25732745](#)]
4. Windsor JW, Kaplan GG. Evolving epidemiology of IBD. Curr Gastroenterol Rep 2019 Jul 23;21(8):40. [doi: [10.1007/s11894-019-0705-6](#)] [Medline: [31338613](#)]
5. Kemp K, Griffiths J, Campbell S, Lovell K. An exploration of the follow-up up needs of patients with inflammatory bowel disease. J Crohns Colitis 2013 Oct;7(9):e386-e395. [doi: [10.1016/j.crohns.2013.03.001](#)] [Medline: [23541150](#)]
6. World Health Organization. Telehealth and Telemedicine Will Henceforth Be Part of the Strategy for Health-for-All. Geneva, Switzerland: World Health Organization; 1997.
7. Siegel CA. Placing value on telemedicine for inflammatory bowel disease. Am J Gastroenterol 2019 Mar 30;114(3):382-383. [doi: [10.14309/ajg.0000000000000151](#)] [Medline: [30768428](#)]
8. Wu C, Wu Z, Yang L, Zhu W, Zhang M, Zhu Q, et al. Evaluation of the clinical outcomes of telehealth for managing diabetes: a PRISMA-compliant meta-analysis. Medicine (Baltimore) 2018 Oct;97(43):e12962 [FREE Full text] [doi: [10.1097/MD.00000000000012962](#)] [Medline: [30412116](#)]
9. Zhao J, Zhai Y, Zhu W, Sun D. Effectiveness of telemedicine for controlling asthma symptoms: a systematic review and meta-analysis. Telemed J E Health 2015 Jun;21(6):484-492. [doi: [10.1089/tmj.2014.0119](#)] [Medline: [25393915](#)]

10. Lundell S, Holmner, Rehn B, Nyberg A, Wadell K. Telehealthcare in COPD: a systematic review and meta-analysis on physical outcomes and dyspnea. *Respir Med* 2015 Jan;109(1):11-26 [FREE Full text] [doi: [10.1016/j.rmed.2014.10.008](https://doi.org/10.1016/j.rmed.2014.10.008)] [Medline: [25464906](https://pubmed.ncbi.nlm.nih.gov/25464906/)]
11. Knox L, Rahman RJ, Beedie C. Quality of life in patients receiving telemedicine enhanced chronic heart failure disease management: A meta-analysis. *J Telemed Telecare* 2017 Aug;23(7):639-649. [doi: [10.1177/1357633X16660418](https://doi.org/10.1177/1357633X16660418)] [Medline: [27450573](https://pubmed.ncbi.nlm.nih.gov/27450573/)]
12. Elkjaer M, Shuhaibar M, Burisch J, Bailey Y, Scherfig H, Laugesen B, et al. E-health empowers patients with ulcerative colitis: a randomised controlled trial of the web-guided 'Constant-care' approach. *Gut* 2010 Dec 11;59(12):1652-1661. [doi: [10.1136/gut.2010.220160](https://doi.org/10.1136/gut.2010.220160)] [Medline: [21071584](https://pubmed.ncbi.nlm.nih.gov/21071584/)]
13. Ozgursoy UBN, Yildirim Y, Senuzun Aykar F, Unsal B. N031 The effect of web-based education on disease activity, symptom management and quality of life in patients with inflammatory bowel disease. *J Crohns Colitis* 2018;12(Suppl 1):S582-S582. [doi: [10.1093/ecco-jcc/jjx180.1045](https://doi.org/10.1093/ecco-jcc/jjx180.1045)]
14. Avery P. Using e-health tools and PROMs to support self-management in patients with inflammatory bowel disease. *Br J Nurs* 2021 Apr 08;30(7):394-402 [FREE Full text] [doi: [10.12968/bjon.2021.30.7.394](https://doi.org/10.12968/bjon.2021.30.7.394)] [Medline: [33830794](https://pubmed.ncbi.nlm.nih.gov/33830794/)]
15. Mao R, Liang J, Wu K, Chen M. Responding to COVID-19: perspectives from the Chinese Society of Gastroenterology. *Gastroenterology* 2020 Jun;158(8):2024-2027 [FREE Full text] [doi: [10.1053/j.gastro.2020.03.046](https://doi.org/10.1053/j.gastro.2020.03.046)] [Medline: [32314563](https://pubmed.ncbi.nlm.nih.gov/32314563/)]
16. de Jong M, van der Meulen-de Jong A, Romberg-Camps M, Degens J, Becx M, Markus T, et al. Development and feasibility study of a telemedicine tool for all patients with IBD: MyIBDcoach. *Inflamm Bowel Dis* 2017 Apr;23(4):485-493. [doi: [10.1097/MIB.0000000000001034](https://doi.org/10.1097/MIB.0000000000001034)] [Medline: [28267047](https://pubmed.ncbi.nlm.nih.gov/28267047/)]
17. Habashi P, Bouchard S, Nguyen GC. Transforming access to specialist care for inflammatory bowel disease: the PACE Telemedicine Program. *J Can Assoc Gastroenterol* 2019 Dec;2(4):186-194 [FREE Full text] [doi: [10.1093/jcag/gwy046](https://doi.org/10.1093/jcag/gwy046)] [Medline: [31616860](https://pubmed.ncbi.nlm.nih.gov/31616860/)]
18. Del Hoyo J, Nos P, Faubel R, Bastida G, Muñoz D, Valero-Pérez E, et al. Adaptation of TECCU app based on patients' perceptions for the telemonitoring of inflammatory bowel disease: a qualitative study using focus groups. *Int J Environ Res Public Health* 2020 Mar 13;17(6):1871 [FREE Full text] [doi: [10.3390/ijerph17061871](https://doi.org/10.3390/ijerph17061871)] [Medline: [32183103](https://pubmed.ncbi.nlm.nih.gov/32183103/)]
19. Atreja A, Khan S, Rogers JD, Ootobo E, Patel NP, Ullman T, HealthPROMISE Consortium Group. Impact of the mobile HealthPROMISE platform on the quality of care and quality of life in patients with inflammatory bowel disease: study protocol of a pragmatic randomized controlled trial. *JMIR Res Protoc* 2015 Feb 18;4(1):e23 [FREE Full text] [doi: [10.2196/resprot.4042](https://doi.org/10.2196/resprot.4042)] [Medline: [25693610](https://pubmed.ncbi.nlm.nih.gov/25693610/)]
20. Singh S, Brill JV, Proudfoot JA, Metcalfe L, Vu L, Sandborn WJ, et al. Project Sonar: a community practice-based intensive medical home for patients with inflammatory bowel diseases. *Clin Gastroenterol Hepatol* 2018 Dec;16(12):1847-1850.e1. [doi: [10.1016/j.cgh.2018.08.052](https://doi.org/10.1016/j.cgh.2018.08.052)] [Medline: [30149146](https://pubmed.ncbi.nlm.nih.gov/30149146/)]
21. Van Deen WK, van der Meulen-de Jong AE, Parekh NK, Kane E, Zand A, DiNicola CA, et al. Development and validation of an inflammatory bowel diseases monitoring index for use with mobile health technologies. *Clin Gastroenterol Hepatol* 2016 Dec;14(12):1742-1750.e7. [doi: [10.1016/j.cgh.2015.10.035](https://doi.org/10.1016/j.cgh.2015.10.035)] [Medline: [26598228](https://pubmed.ncbi.nlm.nih.gov/26598228/)]
22. Louis E, Dotan I, Ghosh S, Mlynarsky L, Reenaers C, Schreiber S. Optimising the inflammatory bowel disease unit to improve quality of care: expert recommendations. *J Crohns Colitis* 2015 Aug;9(8):685-691 [FREE Full text] [doi: [10.1093/ecco-jcc/jjv085](https://doi.org/10.1093/ecco-jcc/jjv085)] [Medline: [25987349](https://pubmed.ncbi.nlm.nih.gov/25987349/)]
23. Ghosh S, Louis E, Beaugier L, Bossuyt P, Bouguen G, Bourreille A, et al. Development of the IBD disk: a visual self-administered tool for assessing disability in inflammatory bowel diseases. *Inflamm Bowel Dis* 2017 Mar;23(3):333-340 [FREE Full text] [doi: [10.1097/MIB.0000000000001033](https://doi.org/10.1097/MIB.0000000000001033)] [Medline: [28146002](https://pubmed.ncbi.nlm.nih.gov/28146002/)]
24. Huang VW, Reich KM, Fedorak RN. Distance management of inflammatory bowel disease: systematic review and meta-analysis. *World J Gastroenterol* 2014 Jan 21;20(3):829-842 [FREE Full text] [doi: [10.3748/wjg.v20.i3.829](https://doi.org/10.3748/wjg.v20.i3.829)] [Medline: [24574756](https://pubmed.ncbi.nlm.nih.gov/24574756/)]
25. Hanlon I, Hewitt C, Bell K, Phillips A, Mikocka-Walus A. Systematic review with meta-analysis: online psychological interventions for mental and physical health outcomes in gastrointestinal disorders including irritable bowel syndrome and inflammatory bowel disease. *Aliment Pharmacol Ther* 2018 Aug 14;48(3):244-259. [doi: [10.1111/apt.14840](https://doi.org/10.1111/apt.14840)] [Medline: [29901820](https://pubmed.ncbi.nlm.nih.gov/29901820/)]
26. Rohde J, Barker JO, Noar SM. Impact of eHealth technologies on patient outcomes: a meta-analysis of chronic gastrointestinal illness interventions. *Transl Behav Med* 2021 Feb 11;11(1):1-10. [doi: [10.1093/tbm/ibz166](https://doi.org/10.1093/tbm/ibz166)] [Medline: [31731292](https://pubmed.ncbi.nlm.nih.gov/31731292/)]
27. George LA, Cross RK. Remote monitoring and telemedicine in IBD: are we there yet? *Curr Gastroenterol Rep* 2020 Feb 10;22(3):12. [doi: [10.1007/s11894-020-0751-0](https://doi.org/10.1007/s11894-020-0751-0)] [Medline: [32040650](https://pubmed.ncbi.nlm.nih.gov/32040650/)]
28. Atanasov P, Gauthier A, Cobos D. The place of eHealth interventions in the management pathway of inflammatory bowel disease (IBD) patients: a systematic review. *Value Health* 2018 Sep;21:S42-S43. [doi: [10.1016/j.jval.2018.07.326](https://doi.org/10.1016/j.jval.2018.07.326)]
29. Cochrane Handbook for Systematic Reviews of Interventions. Cochrane Training. URL: <http://handbook.cochrane.org> [accessed 2020-11-09]
30. Jadad AR, Moore R, Carroll D, Jenkinson C, Reynolds DM, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996 Feb;17(1):1-12. [doi: [10.1016/0197-2456\(95\)00134-4](https://doi.org/10.1016/0197-2456(95)00134-4)] [Medline: [8721797](https://pubmed.ncbi.nlm.nih.gov/8721797/)]

31. Guyatt G, Mitchell A, Irvine E, Singer J, Williams N, Goodacre R, et al. A new measure of health status for clinical trials in inflammatory bowel disease. *Gastroenterology* 1989 Mar;96(3):804-810. [Medline: [2644154](#)]
32. Harvey R, Bradshaw J. A simple index of Crohn's-disease activity. *Lancet* 1980 Mar 08;1(8167):514. [doi: [10.1016/s0140-6736\(80\)92767-1](#)] [Medline: [6102236](#)]
33. Seo M, Okada M, Yao T, Ueki M, Arima S, Okumura M. An index of disease activity in patients with ulcerative colitis. *Am J Gastroenterol* 1992 Aug;87(8):971-976. [Medline: [1642220](#)]
34. Walmsley RS, Ayres RCS, Pounder RE, Allan RN. A simple clinical colitis activity index. *Gut* 1998 Jul;43(1):29-32 [FREE Full text] [doi: [10.1136/gut.43.1.29](#)] [Medline: [9771402](#)]
35. Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. *N Engl J Med* 1987 Dec 24;317(26):1625-1629. [doi: [10.1056/NEJM198712243172603](#)] [Medline: [3317057](#)]
36. Rubin DT, Ananthakrishnan AN, Siegel CA, Sauer BG, Long MD. ACG Clinical Guideline: ulcerative colitis in adults. *Am J Gastroenterol* 2019 Mar;114(3):384-413. [doi: [10.14309/ajg.000000000000152](#)] [Medline: [30840605](#)]
37. Best WR. Predicting the Crohn's disease activity index from the Harvey-Bradshaw Index. *Inflamm Bowel Dis* 2006 Apr;12(4):304-310. [doi: [10.1097/01.MIB.0000215091.77492.2a](#)] [Medline: [16633052](#)]
38. Larsen DL, Attkisson C, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. *Eval Program Plann* 1979;2(3):197-207. [doi: [10.1016/0149-7189\(79\)90094-6](#)] [Medline: [10245370](#)]
39. Cross RK, Cheevers N, Rustgi A, Langenberg P, Finkelstein J. Randomized, controlled trial of home telemanagement in patients with ulcerative colitis (UC HAT). *Inflamm Bowel Dis* 2012 Jun;18(6):1018-1025 [FREE Full text] [doi: [10.1002/ibd.21795](#)] [Medline: [21688350](#)]
40. Cross RK, Langenberg P, Regueiro M, Schwartz DA, Tracy JK, Collins JF, et al. A randomized controlled trial of TELEmedicine for Patients with Inflammatory Bowel Disease (TELE-IBD). *Am J Gastroenterol* 2019 Mar;114(3):472-482. [doi: [10.1038/s41395-018-0272-8](#)] [Medline: [30410041](#)]
41. Schliep M, Chudy-Onwugaje K, Abutaleb A, Langenberg P, Regueiro M, Schwartz DA, et al. TELEmedicine for Patients With Inflammatory Bowel Disease (TELE-IBD) does not improve depressive symptoms or general quality of life compared with standard care at tertiary referral centers. *Crohns Colitis* 360 2020 Jan;2(1):otaa002-otaa008 [FREE Full text] [doi: [10.1093/crocol/otaa002](#)] [Medline: [32201859](#)]
42. Bilgrami Z, Abutaleb A, Chudy-Onwugaje K, Langenberg P, Regueiro M, Schwartz DA, et al. Effect of TELEmedicine for inflammatory bowel disease on patient activation and self-efficacy. *Dig Dis Sci* 2020 Jan 02;65(1):96-103 [FREE Full text] [doi: [10.1007/s10620-018-5433-5](#)] [Medline: [30604373](#)]
43. Miloh T, Shub M, Montes R, Ingebo K, Silber G, Pasternak B. Text messaging effect on adherence in children with inflammatory bowel disease. *J Pediatr Gastroenterol Nutr* 2017 Jun;64(6):939-942. [doi: [10.1097/MPG.0000000000001399](#)] [Medline: [27602705](#)]
44. Hunt M, Rodriguez L, Marcelle E. A cognitive behavioral therapy workbook delivered online with minimal therapist feedback improves quality of life for inflammatory bowel disease patients. *Intern Med J* 2017:1-16.
45. Krier M, Kaltenbach T, McQuaid K, Soetikno R. Potential use of telemedicine to provide outpatient care for inflammatory bowel disease. *Am J Gastroenterol* 2011 Dec;106(12):2063-2067. [doi: [10.1038/ajg.2011.329](#)] [Medline: [22138934](#)]
46. Heida A, Dijkstra A, Muller Kobold A, Rossen JW, Kindermann A, Kokke F, et al. Efficacy of home telemonitoring versus conventional follow-up: a randomized controlled trial among teenagers with inflammatory bowel disease. *J Crohns Colitis* 2018 Mar 28;12(4):432-441. [doi: [10.1093/ecco-jcc/jjx169](#)] [Medline: [29228230](#)]
47. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, Becx MC, Maljaars JP, Cilissen M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. *Lancet* 2017 Sep 02;390(10098):959-968. [doi: [10.1016/S0140-6736\(17\)31327-2](#)] [Medline: [28716313](#)]
48. Linn AJ, van Dijk L, van Weert JC, Gebeyehu BG, van Bodegraven AA, Smit EG. Creating a synergy effect: a cluster randomized controlled trial testing the effect of a tailored multimedia intervention on patient outcomes. *Patient Educ Couns* 2018 Aug;101(8):1419-1426. [doi: [10.1016/j.pec.2018.03.017](#)] [Medline: [29609899](#)]
49. McCombie A, Walmsley R, Barclay M, Ho C, Langlotz T, Regenbrecht H, et al. A noninferiority randomized clinical trial of the use of the smartphone-based health applications IBDsmart and IBDoc in the care of inflammatory bowel disease patients. *Inflamm Bowel Dis* 2020 Jun 18;26(7):1098-1109. [doi: [10.1093/ibd/izz252](#)] [Medline: [31644793](#)]
50. McCombie A, Geary R, Andrews J, Mulder R, Mikocka-Walus A. Does computerized cognitive behavioral therapy help people with inflammatory bowel disease? a randomized controlled trial. *Inflamm Bowel Dis* 2016 Jan;22(1):171-181. [doi: [10.1097/MIB.0000000000000567](#)] [Medline: [26360545](#)]
51. Carlsen K, Jakobsen C, Houen G, Kallemose T, Paerregaard A, Riis LB, et al. Self-managed eHealth disease monitoring in children and adolescents with inflammatory bowel disease: a randomized controlled trial. *Inflamm Bowel Dis* 2017 Mar;23(3):357-365. [doi: [10.1097/MIB.0000000000001026](#)] [Medline: [28221247](#)]
52. Akobeng AK, O'Leary N, Vail A, Brown N, Widiatmoko D, Fagbemi A, et al. Telephone consultation as a substitute for routine out-patient face-to-face consultation for children with inflammatory bowel disease: randomised controlled trial and economic evaluation. *EBioMedicine* 2015 Sep;2(9):1251-1256 [FREE Full text] [doi: [10.1016/j.ebiom.2015.08.011](#)] [Medline: [26501125](#)]

53. Del Hoyo J, Nos P, Faubel R, Muñoz D, Domínguez D, Bastida G, et al. A web-based telemanagement system for improving disease activity and quality of life in patients with complex inflammatory bowel disease: pilot randomized controlled trial. *J Med Internet Res* 2018 Nov 27;20(11):e11602 [FREE Full text] [doi: [10.2196/11602](https://doi.org/10.2196/11602)] [Medline: [30482739](https://pubmed.ncbi.nlm.nih.gov/30482739/)]
54. Vachon A, Scott FI. The treatment approach to inflammatory bowel disease in 2020. *Curr Opin Gastroenterol* 2020 Jul;36(4):247-256. [doi: [10.1097/MOG.0000000000000653](https://doi.org/10.1097/MOG.0000000000000653)] [Medline: [32452863](https://pubmed.ncbi.nlm.nih.gov/32452863/)]
55. Jackson BD, Gray K, Knowles SR, De Cruz P. EHealth technologies in inflammatory bowel disease: a systematic review. *J Crohns Colitis* 2016 Sep;10(9):1103-1121. [doi: [10.1093/ecco-jcc/jjw059](https://doi.org/10.1093/ecco-jcc/jjw059)] [Medline: [26928960](https://pubmed.ncbi.nlm.nih.gov/26928960/)]
56. Shpoliansky M, Shouval DS. [APPROACH TO VERY EARLY ONSET INFLAMMATORY BOWEL DISEASE]. *Harefuah* 2020 Mar;159(3):206-211. [Medline: [32186793](https://pubmed.ncbi.nlm.nih.gov/32186793/)]
57. Topal F, Camyar H, Saritas Yuksel E, Gunay S, Topal F, Gür EÖ. Work productivity loss in inflammatory bowel disease patients in Turkey. *Gastroenterol Res Pract* 2020 Jan 08;2020:6979720-6979726 [FREE Full text] [doi: [10.1155/2020/6979720](https://doi.org/10.1155/2020/6979720)] [Medline: [32382271](https://pubmed.ncbi.nlm.nih.gov/32382271/)]
58. Martelli L, Lopez A, Strobel S, Danese S, Roblin X, Baumann C, et al. Adherence to infliximab therapy in inflammatory bowel disease patients in a real-life setting. *J Dig Dis* 2017 Oct;18(10):566-573. [doi: [10.1111/1751-2980.12539](https://doi.org/10.1111/1751-2980.12539)] [Medline: [28858439](https://pubmed.ncbi.nlm.nih.gov/28858439/)]
59. Kane S, Shaya F. Medication non-adherence is associated with increased medical health care costs. *Dig Dis Sci* 2008 Apr;53(4):1020-1024. [doi: [10.1007/s10620-007-9968-0](https://doi.org/10.1007/s10620-007-9968-0)] [Medline: [17934828](https://pubmed.ncbi.nlm.nih.gov/17934828/)]
60. Hommel KA, Davis CM, Baldassano RN. Medication adherence and quality of life in pediatric inflammatory bowel disease. *J Pediatr Psychol* 2008 Sep;33(8):867-874 [FREE Full text] [doi: [10.1093/jpepsy/jsn022](https://doi.org/10.1093/jpepsy/jsn022)] [Medline: [18337262](https://pubmed.ncbi.nlm.nih.gov/18337262/)]
61. Di Toro R, Lama G. [A case of nephroblastoma associated with the nephrotic syndrome]. *Minerva Pediatr* 1974 Apr 21;26(14):724-727. [Medline: [4376208](https://pubmed.ncbi.nlm.nih.gov/4376208/)]
62. 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](https://doi.org/10.1001/jamainternmed.2015.7667)] [Medline: [26831740](https://pubmed.ncbi.nlm.nih.gov/26831740/)]

Abbreviations

IBD: inflammatory bowel disease

IBDQ: inflammatory bowel disease questionnaire

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

QoL: quality of life

RCT: randomized controlled trial

Edited by R Kukafka, G Eysenbach; submitted 20.03.21; peer-reviewed by T Taft, D López López; comments to author 18.05.21; revised version received 18.06.21; accepted 14.10.21; published 24.03.22.

Please cite as:

Pang L, Liu H, Liu Z, Tan J, Zhou LY, Qiu Y, Lin X, He J, Li X, Lin S, Ghosh S, Mao R, Chen M

Role of Telemedicine in Inflammatory Bowel Disease: Systematic Review and Meta-analysis of Randomized Controlled Trials

J Med Internet Res 2022;24(3):e28978

URL: <https://www.jmir.org/2022/3/e28978>

doi: [10.2196/28978](https://doi.org/10.2196/28978)

PMID: [35323120](https://pubmed.ncbi.nlm.nih.gov/35323120/)

©Lanlan Pang, Hengyu Liu, Zhidong Liu, Jinyu Tan, Long-yuan Zhou, Yun Qiu, Xiaoqing Lin, Jinshen He, Xuehua Li, Sinan Lin, Subrata Ghosh, Ren Mao, Minhu Chen. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 24.03.2022. 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.

Review

Value Cocreation in Health Care: Systematic Review

Yuxin Peng¹, BA; Tailai Wu¹, PhD; Zhuo Chen^{2,3}, PhD; Zhaohua Deng⁴, PhD

¹School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

²Department of Health Policy and Management, University of Georgia College of Public Health, Athens, GA, United States

³School of Economics, Faculty of Humanities and Social Sciences, University of Nottingham Ningbo China, Ningbo, China

⁴School of Management, Huazhong University of Science and Technology, Wuhan, China

Corresponding Author:

Tailai Wu, PhD

School of Medicine and Health Management

Tongji Medical College

Huazhong University of Science and Technology

13 Hangkong Road

Qiaokou District

Wuhan, 430030

China

Phone: 86 134 7707 2665

Email: lncl2012@gmail.com

Abstract

Background: Value cocreation in health care (VCCH), mainly based on service-dominant logic, emphasizes that participants, including both patients and physicians, can effectively enroll in the health care value creation process. Effective VCCH is of great significance for realizing value-based health care and improving doctor-patient relationships. Therefore, a comprehensive understanding of VCCH is critical. However, the current literature on VCCH is fragmented and not well studied.

Objective: The goal of the research is to investigate the antecedents, consequences, and dimensions of VCCH by systematically searching, selecting, summarizing, and evaluating relevant literature.

Methods: English-language articles on VCCH in the Web of Science, PubMed, and Scopus databases published from January 2008 to December 2019 were identified. The articles were screened using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses protocol, and the quality of studies included were appraised using the Mixed Methods Appraisal Tool.

Results: Out of the 181 publications initially identified through the bibliographic searches, 28 publications met the inclusion criteria. This review summarizes antecedents, consequences, and dimensions of VCCH, as well as possible associations among them. An integrative framework is also proposed for mapping the literature of VCCH grounded on social cognitive theory to reveal the whole process of VCCH.

Conclusions: The findings of this systematic review provide implications for continued development of VCCH and contribute to inspire more research in the future.

(*J Med Internet Res* 2022;24(3):e33061) doi:[10.2196/33061](https://doi.org/10.2196/33061)

KEYWORDS

value cocreation; health care; patient value; health care professional value; systematic review

Introduction

Value cocreation, which has received much academic attention in the 21st century, describes cooperation among multiple stakeholders [1,2]. As the main theoretical foundation of value cocreation, service-dominant logic (SDL) suggests that the service provider should not be the only value creator, and service receivers could also cocreate the service value with service provider [3-5]. Since service could be a process of interaction

between an organization and its customers, value of service based on resources used and integrated in the service process is created by the participants together [3]. Therefore, SDL focuses on value from use rather than value from exchange, which shifts the research focus from goods to services. In recent years, value cocreation is applied in the health care area to understand the patient value creation process.

In health care, value cocreation should be emphasized. Value cocreation in health care (VCCH) refers to the integration of

resources through activities and interactions with collaborators to realize the benefit of patients in the health care service delivery network [6]. Therefore, patients and health care providers integrate knowledge, skills, equipment, medicine, facilities, and financial resources to obtain their mutual benefits [7,8]. This definition emphasizes the active participation of patients, who are no longer passive recipients of services but active cocreators. VCCH involves a range of activities around patients or collaboration with service network members, including family members, friends, other patients, health care professionals, and external communities [6]. When patients cocreate health care value with physicians, they participate in the whole processes of treatment. They could offer opinions to improve their compliance [9], eventually improving their health status and service experience [10]. Meanwhile, cocreation between patients and health professionals could also reduce medical costs [11], improve the efficiency of the use of existing medical resources [12,13], and integrate medical resources from different sources to create values with health care professionals [2].

Much of the literature on value cocreation is found in the business field, focusing on the firm and consumer values, yet patient value has attracted attention in the health care field with the changing physician-patient relationship. Patients are increasingly taking an active role in the health care decision-making process [14] and creating value with other participants in the health care delivery process [15]. The relevant notion of value-based health care emphasizes the importance of listening to the patient's voice and advocates providing high value to patients by taking into account their outcomes, needs, and costs when treating their illnesses [16]. However, value-based health care research touches less on the value creation process. Therefore, it is essential to discuss health care value from a value cocreation perspective.

However, for VCCH, the factors are not explored systematically, underlying mechanisms of its factors are vague, and consequences are not fully investigated. To enrich the research related to VCCH and seek solutions to these known issues, we conducted a systematic review to search, select, and evaluate relevant literature, summarize the theories and methods used in existing studies, and explore the antecedents and consequences of VCCH and the dimensions of value cocreation in previous studies to lay a foundation for future studies.

In this study, the existing relevant literature on VCCH was screened following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Multimedia Appendix 1). First, we discuss the development of value cocreation concepts. Second, we describe the process of literature inclusion and selection. Third, we described the available studies, year of publication of the VCCH papers, and the methodological and theoretical underpinnings of the VCCH research. We then summarize the antecedents and consequences

of VCCH and value cocreation dimensions from different settings and actors and propose an integrative framework for understanding value cocreation behavior. Finally, we discuss implications, opportunities for future research, and limitations.

Methods

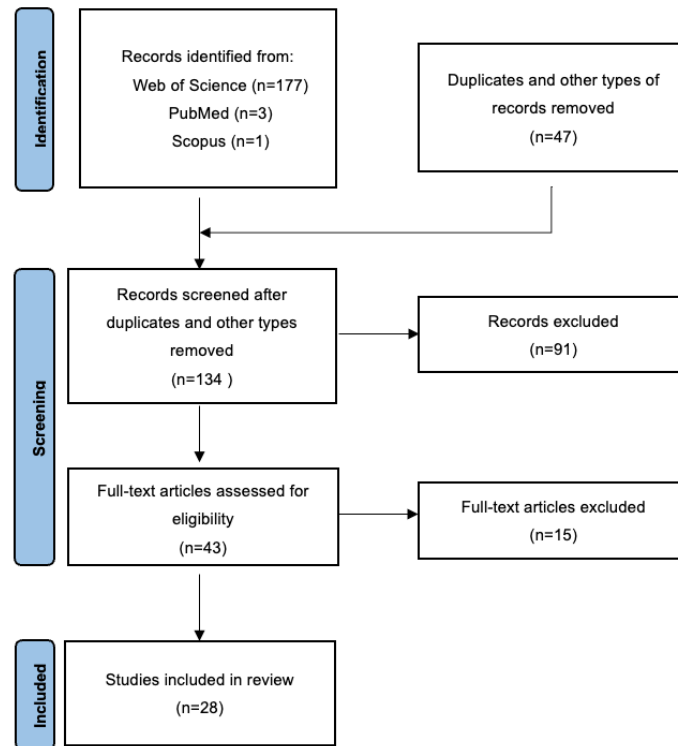
For this study, we used Web of Science as the primary database and PubMed and Scopus as supplementary databases for literature retrieval conducted on May 25, 2020. Keywords, relevant synonyms, and associated truncations used in the search revolved around two concepts—health care area and value cocreation. The search strategy using topic and Boolean/phrase search modes was used to retrieve papers published from 2008 to 2019 and comprised 3 search strings:

1. patient* OR health OR medical OR “health care” OR “healthcare” OR “online health community*” OR eHealth OR E-health OR mHealth OR m-health OR “mobile health”
2. “value co-creat*” OR “value cocreat*” OR “value-co-creat*” OR “co-creat* of value” OR “co-creat* value”
3. 1 and 2

PRISMA framework was used to record the review process. Initial search results yielded 181 papers across the databases. After duplicates and other types of papers (eg, meeting papers, book chapters) were removed, 134 relevant works were screened. Title and abstract screenings were undertaken independently by two trained research assistants, with disagreements about inclusion resolved through discussion with the research coordinator until an agreement was reached. Papers with titles and abstracts not meeting the selection criteria were removed. Inclusion criteria were empirical studies on value cocreation activities between patients and other participants, including doctors, nurses, and other patients, and studies on the value creation process. Exclusion criteria were non-English language papers; meeting papers, books, editorials, news, reports, and patents; unrelated or incomplete papers; and studies not occurring around patients or in the health care domain. A manual search was conducted of each paper's reference list to identify relevant papers not recognized in database searches. Finally, 43 papers remained for full-text review.

The selection criteria were applied to the 43 papers reviewed in full in this paper, and the final number of papers included in this review was 28. A flowchart summarizing the search, screening, and study selection process is shown in Figure 1. We used the Mixed Methods Appraisal Tool (MMAT) developed by Pace et al [17] for appraisal of the quality of qualitative, quantitative, and mixed methods research studies included. The results showed that the quality satisfies the requirements of a systematic review; details can be found in Multimedia Appendix 2.

Figure 1. Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram of the screening process.



Results

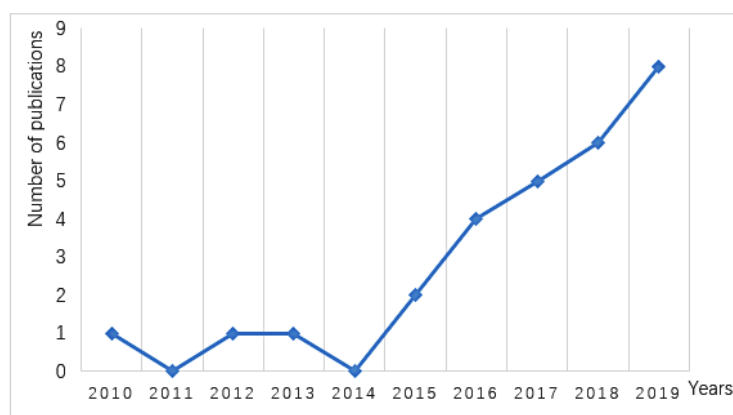
According to our systematic literature review, we received results in 5 main areas: (1) publication years and journals of existing studies, (2) methods used in existing studies, (3) theories applied in existing studies, (4) dimensions of VCCH in existing studies, and (5) antecedents and consequences of VCCH in existing studies.

Publication Years and Journals

As shown in [Figure 2](#), since 2010, there has been a clear upward trend in the research on VCCH. This trend indicates that the application of value cocreation in the medical field is gradually gaining academic attention and is an emerging research area. Following this trend, we expect that more and more studies about VCCH will appear in the coming years.

Toward the publication journals, we found the 28 papers appeared in 17 journals ([Multimedia Appendix 3](#)), with 6 journals publishing more than 2 papers on VCCH, indicating that these journals are more interested in this topic. The most publications are in the Journal of Service Management, with 4 studies related to VCCH. The Journal of Service Theory and Practice, Service Business, and Sustainability has published 3 related studies. As can be seen from this distribution, most current research on VCCH was published in journals in the service management field. Meanwhile, some journals in the public health field have published papers on VCCH, including the International Journal of Environmental Research and Public Health and the International Journal of Pharmaceutical and Healthcare Marketing.

Figure 2. Publication timeline of literature.

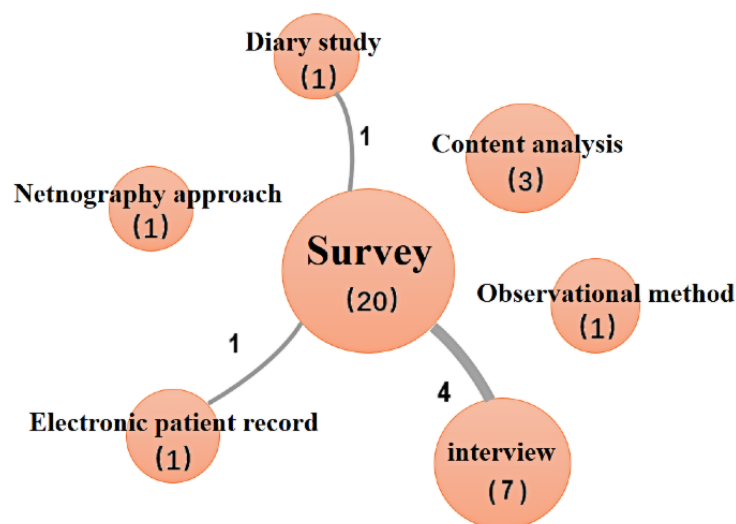


Research Methods

As shown in [Figure 3](#), survey is the research method most used (20/28, 71%). Several studies used survey along with other methods, including interview (depth interview and semistructured depth interview), diary study, and electronic patient record. We found that survey method combined more

with interview method (4 times) than any other methods. This suggests that the interview method allows for a more detailed understanding of participant attitudes and motivations, which facilitates accurate quantitative study. In addition, content analysis, observation, and netnography, an adaptation of ethnography for the contingencies of online community and culture, were used separately in some studies.

Figure 3. Research methods in the literature.



Theories

In the studies included, theories are used in two issues: driving factors of value cocreation behavior/activities and effects of value cocreation behavior/activities ([Multimedia Appendix 4](#)).

On driving factors of value cocreation behavior/activities, patient-related, platform-related, and hospital-related factors are proposed to impact value cocreation. For the driving of patient-related factors, several theories including SDL, consumer culture theory, broaden-and-build theory of positive emotions, social identity theory, self-determination theory, social cognitive theory, practice theory, and construal level theory are applied. For the driving of platform factors, theories like organizational support theory and self-awareness theory are leveraged. For the driving of hospital-related factors, customer training and education perspective are used to understand the effects.

Regarding the effects of value cocreation behavior/activities, self-regulation theory is used to explain the effects of value cocreation on several health outcomes such as well-being, health condition, and health care values.

Dimensions of Value Cocreation in Health Care

In settings where VCCH occurs, different participants are involved in value cocreation with different value creation behaviors. To better understand the dimensions of VCCH, we categorized the literature of VCCH based on the places where VCCH happens and whether VCCH involves the internet. On the places where VCCH happens, since hospitals are the significant medical institutions, we classify VCCH into 2 types based on a typology of VCCH activities: inside-hospital and outside-hospital [18]. With the application of health internet technology, value creators including patients, physicians, and others create values easily and behave differently in offline and

online settings [2,19,20]. Therefore, combining the above 2 categories, 4 dimensions of VCCH are proposed according to whether VCCH happens inside or outside the hospital and offline or online. The dimensions are summarized in [Multimedia Appendix 5](#).

Inside-Hospital Offline Value Cocreation in Health Care

In this dimension, value is cocreated by participants within the hospital and does not involve any health information technology. The main participants include patients, doctors, nurses, and other hospital staff. Several behaviors are discussed in this dimension, including value cocreation behavior, patient participation behavior, patient responsible behavior, and customer effective behavior.

For value cocreation behavior, Yi and Gong [21] identified 2 types: participation behavior and citizenship behavior, while Olsson [22] used complaints and feedback to measure value cocreation behaviors. Regarding patient participation behavior, it is reflected as information sharing [23-26], information seeking [23,26], coproducing, cooperating [6], maintaining interpersonal interaction [6,24,26,27], and enjoying spending time with other patients at the hospital [24], etc. Patient responsible behavior includes following the doctor's instructions to take prescribed medication and complying with the doctor's diagnosis-related recommendations [23,25]. Customer effective behavior includes patients' scientific use of health-related instruments and tracking of disease-related indicators according to the guidance of medical professionals [28].

Inside-Hospital Online Value Cocreation in Health Care

With the application of information technology in health care, health information systems (HIS) including electronic health records are widely used in hospitals to create patient value by health care participants. HIS enables patients to increase their knowledge and connect with their physicians, while HIS makes it easy for health professionals to access information and facilitate communication with patients [29]. Therefore, HIS helps build a network of interdependent health care participants to cocreate their value by facilitating interactions between participants [30].

According to Beirão et al [11], with the support of HIS, VCCH could be divided into 3 levels from the ecosystem perspective: macro, meso, and micro. At the macro level, the main participants are governmental departments, ministries of health, and other organizations that shape national health policies. At this level, the context is the ecosystem within which VCCH happens constantly. At the meso level, the main parties are hospitals, primary care units, and health organizations. At this level, parties serve one another directly and indirectly to cocreate patient value. At the micro level, the main actors are health professionals, patients, and their families. At this level, VCCH generally occurs between actors' dyads.

Outside-Hospital Offline Value Cocreation in Health Care

In this dimension, VCCH happens outside the hospital and does not involve information technologies such as the internet. The main participants in this dimension are patients, families, and their friends. Based on the literature review, VCCH is reflected as patient participation behavior, health-related complementary behavior, and patient self-administration in this dimension.

Patient participation behavior could be indicated as sharing worries and anxieties with others, maintaining good relationships with family and friends [24], and seeking and sharing health information with others [6,31]. Health-related complementary behavior includes monitoring and maintaining a healthy diet, exercising, changing the ways of doing things and interacting with others, and dancing and spending time with children to reduce anxiety [31]. Similar concepts like health behavior changes and complementary therapies [6] have been proposed in previous literature. Health behavior changes contain physical and dietary health behavior changes, while complementary therapies refer to improving one's health status by taking complementary medicines, exercise, yoga, meditation, and other activities [6]. To assess whether innovative therapeutic solutions create value for patients and health systems, Spano et al [32] used patients' self-administration to represent VCCH. Their self-administration is to administer a patient's therapy, such as subcutaneous injection, at home.

Outside-Hospital Online Value Cocreation in Health Care

In this dimension, information technology has been applied outside hospital for health purposes. Many digital apps such as digital health platforms have become popular for patients to seek and use health information. Digital health platforms help

patients connect with others, share information and experiences, and receive support for good health outcomes.

According to Presti et al [33], value cocreation online can be represented as customer engagement, which includes affection, activation, and cognitive dimensions. The affection dimension is the extent to which the patient influences the platform or health care services during the platform interaction; the activation dimension is defined as time, energy, and the energy spent by the patient during the interaction; and the cognitive dimension refers to the cognitive processing activated by the patient during the platform interaction.

From the perspective of the service provider and receiver, 3 types of engagement practices are proposed: information, advising, and empathy practices [34]. In addition, digital information search [35], collation of health information [6,31], and knowledge contribution [36] to other members of the online platform also are part of patient engagement.

Beside customer engagement, connecting with others could also convey VCCH in this dimension [6]. To be specific, patient connection with others could be shown as their experience using online platform [37] and maintaining ongoing relationships among online community members [36].

Antecedents of Value Cocreation in Health Care

To review the antecedents of VCCH, we categorize the previous literature according to 2 dimensions: actors and settings. For actors, as mentioned in the previous section, several actors, including patients, health care professionals, families, and friends, are involved in value cocreation. However, previous literature studies the antecedents mainly from the perspectives of 2 actors: patient and health care professional. For settings, the widespread use of information technologies, especially internet technology in health care, plays an important role in the cocreation of patient value. Therefore, we could classify the antecedents into 4 types: patient offline perspective, patient online perspective, health care professional offline perspective, and health care professional online perspective. The antecedents are all summarized in [Multimedia Appendix 6](#).

Patient Offline Perspective

In this perspective, factors that influence VCCH are studied from the patient perspective in the offline setting. We divide the factors into 2 types according to their attributes: intrinsic factors and extrinsic factors. Intrinsic factors are mainly about patients themselves. For example, positive intrinsic motivation [28,38] and positive emotions [9] are found to facilitate patient participation. Patient empowerment is shown to promote value cocreation behaviors. Patients' trust in physicians [26,39], provider-patient orientation [40], and patient pre-encounter actor value needs [39] affect the cocreation of patient value, which in turn affects the service experience and satisfaction. Finally, personality traits (including agreeableness and extraversion), personal values [10], and gender [22] of patients also have effects on value cocreation behaviors.

Extrinsic factors are factors outside patients. For example, the interaction between patients and physicians, other patients, and their families or friends positively affects value cocreation

behaviors [23,24,31]. Physician performance is found to influence patient information seeking, information sharing, and responsible behavior. Patient perceived transparency is shown to effectively reduce information asymmetry between doctors and patients and influences patients' perceived value and satisfaction [41]. Finally, negative factors including lack of empathy, support, and courtesy from physicians and stereotyping of health professionals can lead to patient dissatisfaction and even complaint behaviors [22,40].

Patient Online Perspective

In this perspective, factors are studied about patients in online settings such as blogs, patient forums, online health communities, and internet hospitals. To be specific, social support, which includes information support, emotional support, and instrumental support, is found to facilitate users' social, functional, and affective values through interaction or cocreation activities among users in online health communities [37,42]. Social identity, which is predicted by integrity trust, benevolence trust, shared vision, and shared language, is found to drive patients' value cocreation activities such as knowledge contribution and continuous willingness to participate [36]. Finally, different types of information processing have also been shown to affect the generation of different kinds of value in online health communities [43]. Meanwhile, users' community experiences and social exclusion have often been used as moderating variables.

Health Care Professional Offline Perspective

From this perspective, factors are studied for health care professionals in offline settings. The main health care professionals involved are doctors and nurses. For doctors, factors determining health care professional behaviors and contributing to value creation include information being provided, patient engagement, trust, physical environment, and collaboration [26]. In addition, a patient's trust in the physician facilitates the physician's management of the patient's condition which facilitates VCCH. For nurses, patient participation, length of stay, and first inpatient stay are found to influence nurses' value creation behaviors such as work engagement, job satisfaction, and helping behaviors [44].

Health Care Professional Online Perspective

In this perspective, factors are studied for health care professionals in online settings. Health information technologies in the health care systems in the world are implemented to address health care challenges, including aging populations living with long-term conditions, persistent variations in the quality of care, and health care cost rising. Health care professionals use of health information technologies in their work can be compulsory or voluntary. The use of health information technologies like electronic health records has becoming a key factor that promotes VCCH by supporting and facilitating multiple interactions among participants at different levels, enabling resource access, sharing, reorganization, and even resource monitoring and institutional generation [11]. Whether a physician is successful in obtaining electronic health information about a patient and the degree of mastery, portability, and credibility of advanced technology will affect

the physician's grasp of patient disease and thus the value cocreation [45].

Consequences of Value Cocreation in Health Care

In VCCH, patients and health care professionals are the main participants, and the cocreated value or other consequences would be mainly distributed between them. Therefore, we could divide the consequences of VCCH into 2 types: patient value and health care professional value. The consequences are summarized in [Multimedia Appendix 7](#).

Patient Value

Patient value is the valuable consequence of value cocreation for patients. Previous literature has discussed many consequences related to health outcomes, service, or overall value. On the health outcomes related consequences, health conditions [26,28,35] and well-being [31,35,40,46] are discussed and found to be helpful in determining the effectiveness of clinical interventions and quality. These consequences can be reflected in patients feeling more energetic, feeling better, and having blood pressure return to normal levels, etc. Moreover, patient compliance, which could predict health outcomes, is also in focus [27]. Finally, health expenditures could reflect patient value according to previous literature.

For service-related consequences, service experience is emphasized as the final goal of value creation [11]. Meanwhile, perceived service quality [9,24,41] and patient satisfaction [24,39-41,47,48], positive word-of-mouth, customer loyalty [10,24], and service engagement [26,27] are also considered as important service-related consequences of VCCH.

Finally, some studies investigate patient value directly. For example, cure and care values are shown to be created by patients through posting and communications with others in online health communities [43]. Perceived value, which includes perceived benefits and perceived cost, could be the result of a patient experience situation in the service contact [37] or patient participation in services [38,49]. Perceived value is also divided into process value and outcome value [23].

Health Care Professional Value

The main health professionals studied are doctors and nurses. For the value of doctors, effectiveness and efficiency of their work can be achieved by effective value cocreation activities [11]. Work effectiveness includes higher diagnostic accuracy, prescription accuracy and safety, prescription compliance, and higher patient orientation while work efficiency means fewer repetitive tests, shorter consultation times, faster work procedures, and avoidance of lost documents. For nurses, the value they created could be predicted by patients' experience of their hospital admission and length of stay, reflected by job satisfaction and technical skills among the nurses [44].

Integrative Framework for Value Cocreation in Health Care

To provide a holistic view of VCCH, we propose an integrative framework by summarizing all of the reviews based on social cognitive theory ([Figure 4](#)). The proposed integrative framework could have sufficient impact on journal publication in several ways.

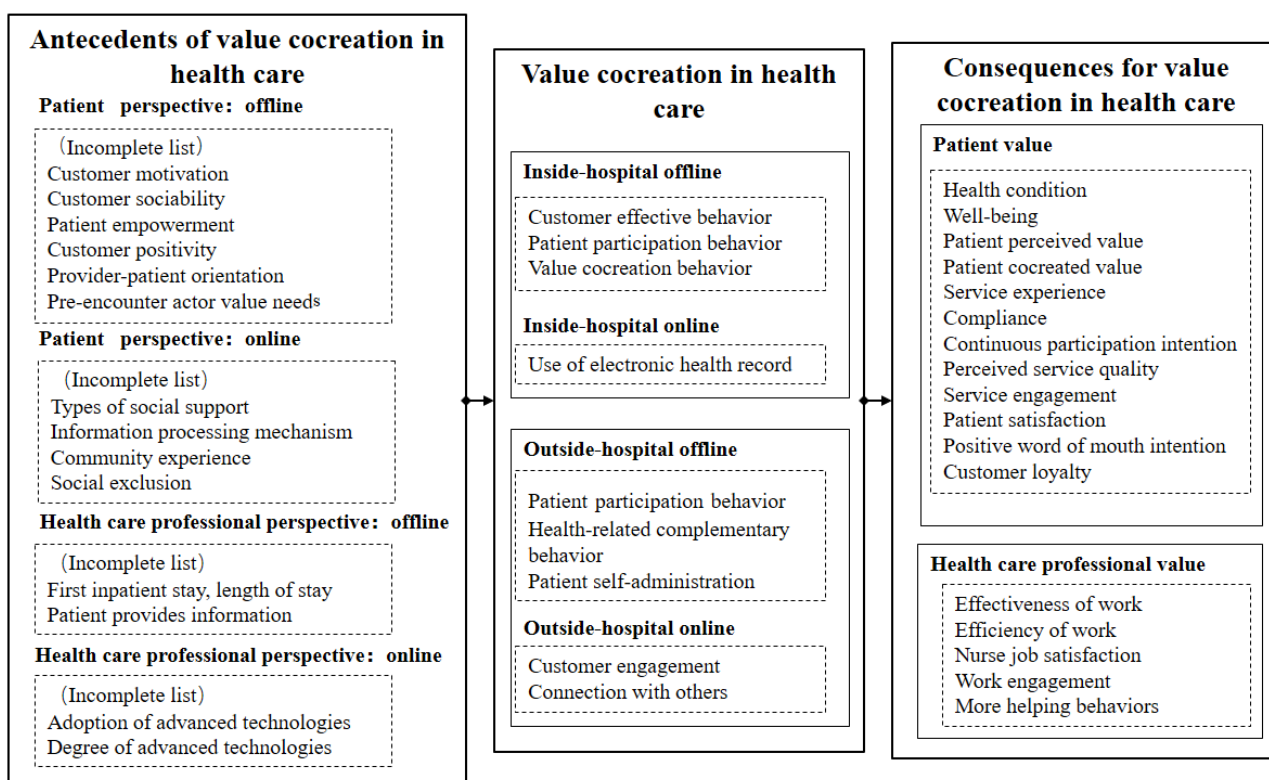
First, the factors and their relationships related to VCCH are systematically mapped and visualized in the framework. In the integrative framework of VCCH, the relationships among antecedents, consequences, and dimensions of VCCH are revealed. This framework could then be used to explain, predict, and evaluate VCCH systematically and comprehensively. Meanwhile, the framework of VCCH also makes the formation process of values for patients and health care professionals through their cocreation transparent. Therefore, this framework helps open the black box of VCCH.

Second, the framework provides a novel theoretical perspective for VCCH. In previous literature, SDL serves as the dominant theoretical perspective for VCCH. However, this framework is proposed based on social cognitive theory, which could bring

a novel theoretical perspective for VCCH. Social cognitive theory proposes that people’s behaviors results from learning in social context, while environmental and personal factors would determine the whole learning process. Therefore, VCCH behaviors could be treated as learned behaviors and would be determined by environmental and personal factors.

Finally, the framework implies many future research directions. Based on the research gaps and relationships conveyed by the framework, many future research directions exist, such as alternative research methods and applications in different contexts. VCCH in online settings or mixed settings should be further investigated, and other types of values creation should be considered.

Figure 4. Framework of value cocreation in health care based on social cognitive theory.



Discussion

Principal Findings

The purpose of this study was to conduct a systematic review of the literature on VCCH. In this review, we focused on empirical studies about VCCH. After a comprehensive search of the databases, 28 journal articles were identified by rigorous application of inclusion and exclusion criteria.

We present our review results from 6 aspects: publication years and journals, research methods, theories, dimensions of VCCH, antecedents of VCCH, and consequences of VCCH. To be specific, we depict the research methods and theories used in VCCH literature. Meanwhile, we categorize the VCCH, its antecedents, and consequences into several dimensions according to 2 main criteria: actors and settings. Actors are mainly patients and health care professionals, while settings include hospitals and the internet. Finally, to summarize our

literature review systematically and provide theoretical insights, we propose an integrative framework to map the literature grounded on social cognitive theory.

Implications

We believe that the findings of this study could convey several implications for both theories and practice. For theoretical implications, this study provides a comprehensive and systematic literature review of VCCH. With the application of SDL in health care, studies about value cocreation are emerging. However, none of the previous literature has provided a systematic summary to reflect the trends, gaps, and directions of VCCH research from different perspectives. Thus, we conducted a systematic review of literature on VCCH. Based on our review, we presented the trends of publication, research methods, and theories of VCCH. Meanwhile, the gaps and research directions of VCCH could be analyzed and speculated about according to our proposed dimensions, antecedents, and

consequences of VCCH. We not only summarize previous literature but also propose an integrative framework of VCCH to map previous literature and guide future research. We integrate the dimensions, antecedents, and consequences of VCCH into the integrative framework to reveal the relationships among the 3 aspects of VCCH from a social cognitive theory perspective.

For practical implications, our systematic review confirms the effectiveness of VCCH. Our study provides a higher quality of evidence compared with single studies about VCCH. Policy makers and health care practitioners can encourage patients and health care professionals to create values together. The uncovered dimensions of VCCH could be the objectives of value cocreation activities. Our summarized dimensions provide a systematic map of VCCH to guide value cocreation practices and improve the feasibility of policies about VCCH. Finally, the antecedents could be the predictors of VCCH. Thus, policy makers and health care practitioners could take action according to antecedents to promote VCCH.

Future Research Opportunities

Our systematic literature review enables us to highlight some opportunities for future research. First, other research methods could be used to conduct more rigorous studies of VCCH. Different methods have advantages and disadvantages in investigating VCCH. Although questionnaires and surveys have dominated previous literature, other methods including randomized controlled trial (RCT) or experiment, case study, or secondary data analysis may have more advantages [50]. For example, compared with surveys, experiments and RCTs have advantages in internal validity that provide strong evidence of causality by minimizing various possible biases, balancing out confounding factors, and improving the validity of statistical tests. Meanwhile, a case study could provide comprehensive and rich information of research objects and inspire unique insights. Finally, secondary data analysis allows investigation of the dynamics of research objects. Therefore, other research methods could be used to better understand VCCH. Because of the complexity of related questions, methods could be used simultaneously rather than by themselves [51].

Second, contextualized and special developed theories are needed for VCCH. Theories could help solve research questions by providing support for analysis, explaining and predicting the research questions [52]. Contextualized and special developed theories could better solve questions related to VCCH compared with theories borrowed from other contexts or domains since contextualized or special developed theories could better reflect the context and regulation of VCCH [53]. The reflection of context of VCCH could capture the person-situation interaction, linkage between personal activities and value creation, and variation among different constructs and convey the application feasibility of study of VCCH [54]. In VCCH, although SDL is found to be the main theoretical foundation in previous literature, contextualization of SDL or building special theories promotes the understanding of VCCH.

Third, VCCH in online settings or mixed settings should be further investigated. With the application of emerging information technologies in health care including artificial

intelligence, big data, and cloud computing, creation of patient value has been changed [55]. Many health care activities happen not only in offline settings but also in online settings or other settings like telemedicine, eHealth, or mobile health in recent years. For example, patients and physicians interact with each other in online health communities to create value by solving the health concerns of patients and rewarding physicians with money or reputation [56]. Meanwhile, electronic health records facilitate the cocreation of value by allowing patients to record their health information and physicians to know patients' information comprehensively. Finally, telemedicine help create value by letting patients and physicians contact each other remotely. Characteristics of different settings would bring different opportunities and challenges to health care and change value cocreation. Therefore, it will be necessary to investigate VCCH in online or mixed settings.

Fourth, more types of values should be considered. Based on our review, values for patients, doctors, and nurses are explored in previous literature. However, more actors have been participating in value cocreation, like friends, families, peers, and allied health professionals, and what they value should be different from patients and doctors; few studies in the previous literature concern their values in the cocreation process. In addition, activities and interactions between doctors and patients could translate into value of the health care organization or even the whole health care system. Values and value cocreation for health professionals and patients in primary, secondary, and tertiary levels of care could be different and are worthy of investigation in the future. Therefore, no matter the value of the patient, health care professional, or individual, value at the organizational or health care system level should also be regarded to better reflect the effectiveness of VCCH.

Finally, since we map previous literature onto social cognitive theory, many related factors could be considered in future study. According to social cognitive theory, factors could be divided into 5 categories: outcome expectancies, social learning, self-efficacy, self-regulation, moral engagement, and other environmental factors [57]. All factors are important antecedents of people's value cocreation behaviors. All the factors are important antecedents of people's value cocreation behaviors. Outcome expectancy is defined as an expectation that an outcome will follow a given behavior [58], while social learning is learning from the knowledge and experience of others we know and trust [59]. Self-efficacy is defined as beliefs about one's ability to perform a specific behavior [60], while self-regulation refers to self-generated thoughts, affects, and behaviors that are systematically oriented toward attainment of one's goals [61]. On environmental factors, they could contain social networks, moral engagement, and culture, etc. Therefore, we would check previously studied factors and see whether any factors from these 5 categories are unexplored and feasible to be studied.

Limitations

There are limitations of this systematic review. Our results are based on journal articles that fit our inclusion criteria and do not include other types of literature (eg, conference proceedings), which may miss some recent research. Thus, the

search scope of the literature could be further expanded. Another limitation is that this study primarily screened for studies strictly related to VCCH. Other types of studies like conceptual research or papers that focus on other actors than patients in VCCH may also provide some insights.

Conclusion

This study presents findings and implications of a systematic review of 28 journal articles for value cocreation in health care.

Our review findings are presented as publication years and journals, research methods, theories, dimensions of VCCH, antecedents of VCCH, and consequences of VCCH. Based on review findings, we proposed an integrative framework based on social cognitive theory to systematically map the literature and reveal the whole value cocreation process in health care. We believe that our literature review and theoretical framework could contribute to the deep understanding of VCCH and inspire more research in the future.

Acknowledgments

This work was supported by grant 71801100 from the National Natural Science Foundation of China, grant 2019kfyXJJS172 from the Fundamental Research Funds for the Central Universities, and grant 71971092 from the National Natural Science Foundation of China.

Authors' Contributions

YP wrote the original draft and curated the data. TW reviewed and edited the manuscript, supervised the project, and created the methodology. ZC reviewed and edited the manuscript. ZD conceptualized and supervised the project.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Preferred Reporting Items for Systematic Reviews and Meta-Analyses checklist.

[[DOCX File , 31 KB - jmir_v24i3e33061_app1.docx](#)]

Multimedia Appendix 2

Quality assessment of included studies.

[[DOCX File , 20 KB - jmir_v24i3e33061_app2.docx](#)]

Multimedia Appendix 3

List of published journals.

[[DOCX File , 14 KB - jmir_v24i3e33061_app3.docx](#)]

Multimedia Appendix 4

Theoretical foundations in the literature.

[[DOCX File , 14 KB - jmir_v24i3e33061_app4.docx](#)]

Multimedia Appendix 5

Dimensions of value cocreation in health care.

[[DOCX File , 19 KB - jmir_v24i3e33061_app5.docx](#)]

Multimedia Appendix 6

Antecedents of value cocreation in health care.

[[DOCX File , 17 KB - jmir_v24i3e33061_app6.docx](#)]

Multimedia Appendix 7

Consequences of value cocreation in health care.

[[DOCX File , 17 KB - jmir_v24i3e33061_app7.docx](#)]

References

1. Prahalad C, Ramaswamy V. Co-creation experiences: the next practice in value creation. *J Interact Mark* 2004 Jan;18(3):5-14. [doi: [10.1002/dir.20015](https://doi.org/10.1002/dir.20015)]
2. Prahalad C, Ramaswamy V. Co-creating unique value with customers. *Strategy Leadership* 2004 Jun;32(3):4-9. [doi: [10.1108/10878570410699249](https://doi.org/10.1108/10878570410699249)]

3. Vargo SL, Lusch RF. Evolving to a new dominant logic for marketing. *J Mark* 2004 Jan;68(1):1-17. [doi: [10.1509/jmkg.68.1.1.24036](https://doi.org/10.1509/jmkg.68.1.1.24036)]
4. Vargo SL, Maglio PP, Akaka MA. On value and value co-creation: a service systems and service logic perspective. *Eur Manag J* 2008 Jun;26(3):145-152. [doi: [10.1016/j.emj.2008.04.003](https://doi.org/10.1016/j.emj.2008.04.003)]
5. Vargo SL, Lusch RF. "Relationship" in transition: an introduction to the special issue on relationship and service-dominant logic. *J Bus Mark Manag* 2010 Oct 23;4(4):167-168. [doi: [10.1007/s12087-010-0042-4](https://doi.org/10.1007/s12087-010-0042-4)]
6. McColl-Kennedy JR, Vargo SL, Dagger TS, Sweeney JC, Kasteren YV. Health care customer value cocreation practice styles. *J Serv Res* 2012 May 1;15(4):370-389. [doi: [10.1177/1094670512442806](https://doi.org/10.1177/1094670512442806)]
7. Vargo SL, Akaka MA. Service-dominant logic as a foundation for service science: clarifications. *Serv Sci* 2009;1(1):32-41. [doi: [10.1287/serv.1.1.32](https://doi.org/10.1287/serv.1.1.32)]
8. Vargo SL, Lusch RF. Service-dominant logic: continuing the evolution. *J Acad Mark Sci* 2007 Aug 1;36(1):1-10. [doi: [10.1007/s11747-007-0069-6](https://doi.org/10.1007/s11747-007-0069-6)]
9. Gallan AS, Jarvis CB, Brown SW, Bitner MJ. Customer positivity and participation in services: an empirical test in a health care context. *J Acad Mark Sci* 2012 Jul 13;41(3):338-356. [doi: [10.1007/s11747-012-0307-4](https://doi.org/10.1007/s11747-012-0307-4)]
10. Ferguson RJ, Paulin M, Bergeron J. Customer sociability and the total service experience. *J Serv Manag* 2010;21(1):25-44. [doi: [10.1108/09564231011025100](https://doi.org/10.1108/09564231011025100)]
11. Beirão G, Patrício L, Fisk RP. Value cocreation in service ecosystems. *J Serv Manag* 2017;28(2):227-249. [doi: [10.1108/josm-11-2015-0357](https://doi.org/10.1108/josm-11-2015-0357)]
12. Ostrom AL, Bitner MJ, Brown SW, Burkhard KA, Goul M, Smith-Daniels V, et al. Moving forward and making a difference: research priorities for the science of service. *J Serv Res* 2010 Jan 18;13(1):4-36. [doi: [10.1177/1094670509357611](https://doi.org/10.1177/1094670509357611)]
13. Elwyn G, Durand MA, Song J, Aarts J, Barr PJ, Berger Z, et al. A three-talk model for shared decision making: multistage consultation process. *BMJ* 2017 Nov 06;359:j4891 [FREE Full text] [doi: [10.1136/bmj.j4891](https://doi.org/10.1136/bmj.j4891)] [Medline: [29109079](https://pubmed.ncbi.nlm.nih.gov/29109079/)]
14. Damman OC, Jani A, de Jong B, Becker A, Metz MJ, de Bruijne M, et al. The use of PROMs and shared decision-making in medical encounters with patients: an opportunity to deliver value-based health care to patients. *J Eval Clin Pract* 2020 Apr;26(2):524-540 [FREE Full text] [doi: [10.1111/jep.13321](https://doi.org/10.1111/jep.13321)] [Medline: [31840346](https://pubmed.ncbi.nlm.nih.gov/31840346/)]
15. 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](https://doi.org/10.1186/s12913-018-3818-y)] [Medline: [30621688](https://pubmed.ncbi.nlm.nih.gov/30621688/)]
16. Porter ME. Value-based health care delivery. *Ann Surg* 2008 Oct;248(4):503-509. [doi: [10.1097/sla.0b013e31818a43af](https://doi.org/10.1097/sla.0b013e31818a43af)]
17. Pace R, Pluye P, Bartlett G, Macaulay AC, Salsberg J, Jagosh J, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. *Int J Nurs Stud* 2012 Jan;49(1):47-53. [doi: [10.1016/j.ijnurstu.2011.07.002](https://doi.org/10.1016/j.ijnurstu.2011.07.002)] [Medline: [21835406](https://pubmed.ncbi.nlm.nih.gov/21835406/)]
18. Sweeney JC, Danaher TS, McColl-Kennedy JR. Customer effort in value cocreation activities. *J Serv Res* 2015;18(3):318-335. [doi: [10.1177/1094670515572128](https://doi.org/10.1177/1094670515572128)]
19. McColl-Kennedy JR, Vargo SL, Dagger TS, Sweeney JC, Kasteren YV. Health care customer value cocreation practice styles. *J Serv Res* 2012 May 1;15(4):370-389. [doi: [10.1177/1094670512442806](https://doi.org/10.1177/1094670512442806)]
20. Nambisan P, Nambisan S. Models of consumer value cocreation in health care. *Health Care Manag Rev* 2009;34(4):344-354. [doi: [10.1097/HMR.0b013e3181abd528](https://doi.org/10.1097/HMR.0b013e3181abd528)] [Medline: [19858919](https://pubmed.ncbi.nlm.nih.gov/19858919/)]
21. Yi Y, Gong T. Customer value co-creation behavior: scale development and validation. *J Bus Res* 2013 Sep;66(9):1279-1284. [doi: [10.1016/j.jbusres.2012.02.026](https://doi.org/10.1016/j.jbusres.2012.02.026)]
22. Olsson EM. Interpersonal complaints regarding cancer care through a gender lens. *Int J Health Care Qual Assur* 2016 Jul 11;29(6):687-702. [doi: [10.1108/IJHCQA-03-2014-0032](https://doi.org/10.1108/IJHCQA-03-2014-0032)] [Medline: [27298065](https://pubmed.ncbi.nlm.nih.gov/27298065/)]
23. Hau LN, Tram Anh PN, Thuy PN. The effects of interaction behaviors of service frontliners on customer participation in the value co-creation: a study of health care service. *Serv Bus* 2016 Mar 16;11(2):253-277. [doi: [10.1007/s11628-016-0307-4](https://doi.org/10.1007/s11628-016-0307-4)]
24. Kim J. Customers' value co-creation with healthcare service network partners. *J Serv Theory Pract* 2019 Oct 02;29(3):309-328. [doi: [10.1108/jstp-08-2018-0178](https://doi.org/10.1108/jstp-08-2018-0178)]
25. Nguyen Hau L, Thuy PN. Customer participation to co-create value in human transformative services: a study of higher education and health care services. *Serv Bus* 2015 Jul 12;10(3):603-628. [doi: [10.1007/s11628-015-0285-y](https://doi.org/10.1007/s11628-015-0285-y)]
26. Osei-Frimpong K, Owusu-Frimpong N. Value co-creation in health care: a phenomenological examination of the doctor-patient encounter. *J Nonprofit Public Sector Market* 2017 Jul 18;29(4):365-384. [doi: [10.1080/10495142.2017.1326356](https://doi.org/10.1080/10495142.2017.1326356)]
27. Osei-Frimpong K, Wilson A, Lemke F. Patient co-creation activities in healthcare service delivery at the micro level: the influence of online access to healthcare information. *Technol Forecasting Soc Change* 2018 Jan;126:14-27. [doi: [10.1016/j.techfore.2016.04.009](https://doi.org/10.1016/j.techfore.2016.04.009)]
28. Damali U, Miller JL, Fredendall LD, Moore D, Dye CJ. Co-creating value using customer training and education in a healthcare service design. *J Op Manag* 2016 Nov 30;47-48(1):80-97. [doi: [10.1016/j.jom.2016.10.001](https://doi.org/10.1016/j.jom.2016.10.001)]
29. Pinho N, Beirao G, Patricio L, Fisk R. Understanding value co-creation in complex services with many actors. *J Serv Manag* 2014;25(4):2014-2093. [doi: [10.1108/josm-02-2014-0055](https://doi.org/10.1108/josm-02-2014-0055)]

30. Thorbjørnsen H, Pedersen PE, Nysveen H. Categorizing networked services. *Eur J Mark* 2009 Apr 03;43(3/4):371-397. [doi: [10.1108/03090560910935488](https://doi.org/10.1108/03090560910935488)]
31. McColl-Kennedy JR, Hogan SJ, Witell L, Snyder H. Cocreative customer practices: effects of health care customer value cocreation practices on well-being. *J Bus Res* 2017 Jan;70:55-66. [doi: [10.1016/j.jbusres.2016.07.006](https://doi.org/10.1016/j.jbusres.2016.07.006)]
32. Spanò R, Di Paola N, Bova M, Barbarino A. Value co-creation in healthcare: evidence from innovative therapeutic alternatives for hereditary angioedema. *BMC Health Serv Res* 2018 Jul 20;18(1):571 [FREE Full text] [doi: [10.1186/s12913-018-3389-y](https://doi.org/10.1186/s12913-018-3389-y)] [Medline: [30029666](https://pubmed.ncbi.nlm.nih.gov/30029666/)]
33. Lo Presti L, Testa M, Marino V, Singer P. Engagement in healthcare systems: adopting digital tools for a sustainable approach. *Sustainability* 2019 Jan 04;11(1):220. [doi: [10.3390/su11010220](https://doi.org/10.3390/su11010220)]
34. Stadtelmann M, Woratschek H, Diederich C. Actor engagement in online health communities. *Int J Pharm Healthc Mark* 2019 Nov 04;13(4):500-515. [doi: [10.1108/ijphm-06-2018-0033](https://doi.org/10.1108/ijphm-06-2018-0033)]
35. Dahl AJ, Peltier JW, Milne GR. Development of a value co-creation wellness model: the role of physicians and digital information seeking on health behaviors and health outcomes. *J Consum Aff* 2018 Jan 18;52(3):562-594. [doi: [10.1111/joca.12176](https://doi.org/10.1111/joca.12176)]
36. Zhao J, Wang T, Fan X. Patient value co-creation in online health communities. *J Serv Manag* 2015 Mar 16;26(1):72-96. [doi: [10.1108/josm-12-2013-0344](https://doi.org/10.1108/josm-12-2013-0344)]
37. Stewart Loane S, Webster CM, D'Alessandro S. Identifying consumer value co-created through social support within online health communities. *J Macromark* 2014 May 29;35(3):353-367. [doi: [10.1177/0276146714538055](https://doi.org/10.1177/0276146714538055)]
38. Osei-Frimpong K. Patient participatory behaviours in healthcare service delivery. *J Serv Theory Pract* 2017;27(2):453-474. [doi: [10.1108/jstp-02-2016-0038](https://doi.org/10.1108/jstp-02-2016-0038)]
39. Osei-Frimpong K. Examining the effects of patient characteristics and prior value needs on the patient-doctor encounter process in healthcare service delivery. *Int J Pharm Healthc Mark* 2016 Jun 06;10(2):192-213. [doi: [10.1108/ijphm-01-2016-0005](https://doi.org/10.1108/ijphm-01-2016-0005)]
40. Huetten ASJ, Antons D, Breidbach C, Piening EP, Salge TO. The impact of occupational stereotypes in human-centered service systems. *J Serv Manag* 2019 Jan 14;30(1):132-155. [doi: [10.1108/josm-12-2016-0324](https://doi.org/10.1108/josm-12-2016-0324)]
41. Yang Y. Is transparency a double-edged sword in citizen satisfaction with public service? Evidence from China's public healthcare. *J Serv Theory Pract* 2018 Jun 11;28(4):484-506. [doi: [10.1108/jstp-06-2017-0102](https://doi.org/10.1108/jstp-06-2017-0102)]
42. Liu W, Fan X, Ji R, Jiang Y. Perceived community support, users' interactions, and value co-creation in online health community: the moderating effect of social exclusion. *Int J Environ Res Public Health* 2019 Dec 27;17(1):1 [FREE Full text] [doi: [10.3390/ijerph17010204](https://doi.org/10.3390/ijerph17010204)] [Medline: [31892188](https://pubmed.ncbi.nlm.nih.gov/31892188/)]
43. Van Oerle S, Lievens A, Mahr D. Value co-creation in online healthcare communities: the impact of patients' reference frames on cure and care. *Psychol Mark* 2018 May 23;35(9):629-639. [doi: [10.1002/mar.21111](https://doi.org/10.1002/mar.21111)]
44. Ding B, Liu W, Tsai S, Gu D, Bian F, Shao X. Effect of patient participation on nurse and patient outcomes in inpatient healthcare. *Int J Environ Res Public Health* 2019 Apr 15;16(8):1 [FREE Full text] [doi: [10.3390/ijerph16081344](https://doi.org/10.3390/ijerph16081344)] [Medline: [30991640](https://pubmed.ncbi.nlm.nih.gov/30991640/)]
45. Lee D. Effects of key value co-creation elements in the healthcare system: focusing on technology applications. *Serv Bus* 2018 Oct 9;13(2):389-417. [doi: [10.1007/s11628-018-00388-9](https://doi.org/10.1007/s11628-018-00388-9)]
46. Kim J. The effect of patient participation through physician's resources on experience and wellbeing. *Sustainability* 2018 Jun 20;10(6):2102. [doi: [10.3390/su10062102](https://doi.org/10.3390/su10062102)]
47. Grönroos C. Service logic revisited: who creates value? And who co-creates? *Eur Bus Rev* 2008 Jun 27;20(4):298-314. [doi: [10.1108/09555340810886585](https://doi.org/10.1108/09555340810886585)]
48. Russo G, Moretta Tartaglione A, Cavacece Y. Empowering patients to co-create a sustainable healthcare value. *Sustainability* 2019 Mar 02;11(5):1315. [doi: [10.3390/su11051315](https://doi.org/10.3390/su11051315)]
49. Zeithaml VA. Consumer perceptions of price, quality, and value: a means-end model and synthesis of evidence. *J Mark* 2018 Nov 19;52(3):2-22. [doi: [10.1177/002224298805200302](https://doi.org/10.1177/002224298805200302)]
50. Bhattacharjee A. *Social Science Research: Principles, Methods, and Practices*. Tampa: Open University Press; 2012.
51. Venkatesh V, Brown SA, Bala H. Bridging the qualitative-quantitative divide: guidelines for conducting mixed methods research in information systems. *Mis Q* 2013 Jan 1;37(1):21-54. [doi: [10.25300/misq/2013/37.1.02](https://doi.org/10.25300/misq/2013/37.1.02)]
52. Glanz K, Bishop DB. The role of behavioral science theory in development and implementation of public health interventions. *Ann Rev Public Health* 2010 Mar 01;31(1):399-418. [doi: [10.1146/annurev.publhealth.012809.103604](https://doi.org/10.1146/annurev.publhealth.012809.103604)]
53. Bamberger P. From the editors beyond contextualization: using context theories to narrow the micro-macro gap in management research. *Acad Manag J* 2008 Oct;51(5):839-846. [doi: [10.5465/amj.2008.34789630](https://doi.org/10.5465/amj.2008.34789630)]
54. Johns G. The essential impact of context on organizational behavior. *Acad Manag Rev* 2006 Apr;31(2):386-408. [doi: [10.5465/amr.2006.20208687](https://doi.org/10.5465/amr.2006.20208687)]
55. Yang J, Li J, Mulder J, Wang Y, Chen S, Wu H, et al. Emerging information technologies for enhanced healthcare. *Comput Industry* 2015 May;69:3-11. [doi: [10.1016/j.compind.2015.01.012](https://doi.org/10.1016/j.compind.2015.01.012)]
56. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. *J Med Internet Res* 2008;10(3):e15 [FREE Full text] [doi: [10.2196/jmir.1053](https://doi.org/10.2196/jmir.1053)] [Medline: [18504244](https://pubmed.ncbi.nlm.nih.gov/18504244/)]

57. Glanz K, Rimer B. Health Behavior and Health Education: Theory, Research, and Practice. Hoboken: John Wiley & Sons; 2008.
58. Bandura A. Social Foundations of Thought and Action: A Socio-Cognitive Theory. Englewood Cliffs: Prentice Hall; 1986.
59. Chen A, Lu Y, Wang B. Customers' purchase decision-making process in social commerce: a social learning perspective. *Int J Inf Manag* 2017 Dec;37(6):627-638. [doi: [10.1016/j.ijinfomgt.2017.05.001](https://doi.org/10.1016/j.ijinfomgt.2017.05.001)]
60. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev* 1977;84(2):191-215. [doi: [10.1037/0033-295X.84.2.191](https://doi.org/10.1037/0033-295X.84.2.191)]
61. Zimmerman B. Chapter 2: Attaining self-regulation: a social cognitive perspective. In: *Handbook of Self-Regulation*. San Diego: Academic Press; 2000:13-39.

Abbreviations

HIS: health information systems

MMAT: Mixed Methods Appraisal Tool

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

RCT: randomized controlled trial

SDL: service-dominant logic

VCCH: value cocreation in health care

Edited by A Mavragani; submitted 23.08.21; peer-reviewed by J Zhou, R Spencer; comments to author 20.01.22; revised version received 25.02.22; accepted 15.03.22; published 25.03.22.

Please cite as:

Peng Y, Wu T, Chen Z, Deng Z

Value Cocreation in Health Care: Systematic Review

J Med Internet Res 2022;24(3):e33061

URL: <https://www.jmir.org/2022/3/e33061>

doi: [10.2196/33061](https://doi.org/10.2196/33061)

PMID: [35333183](https://pubmed.ncbi.nlm.nih.gov/35333183/)

©Yuxin Peng, Tailai Wu, Zhuo Chen, Zhaohua Deng. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 25.03.2022. 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.

Review

Telemedicine Acceptance Among Older Adult Patients With Cancer: Scoping Review

Ning-Qi Pang^{1*}, MBBS, MMed, MPH; Jerrald Lau^{2,3*}, BPsy, MPH; Si-Ying Fong²; Celine Yu-Hui Wong², BSc; Ker-Kan Tan^{2,4}, MBBS, MMed, PhD

¹Division of Hepatobiliary & Pancreatic Surgery and Liver Transplantation, National University of Singapore, Singapore, Singapore

²Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

³Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

⁴Division of Colorectal Surgery, University Surgical Cluster, National University Hospital, Singapore, Singapore

*these authors contributed equally

Corresponding Author:

Ker-Kan Tan, MBBS, MMed, PhD

Yong Loo Lin School of Medicine

National University of Singapore

NUHS Tower Block, Level 8

1E Kent Ridge Road

Singapore, 119228

Singapore

Phone: 65 67724220

Email: surtkk@nus.edu.sg

Abstract

Background: Cancer is likely to remain the most prevalent noncommunicable disease in high-income countries with an older population. Interestingly, no review of attitudes toward telemedicine among older adults has been performed. This is likely to be the group most affected by both cancer and the increasing use of technology in health care.

Objective: We aimed to map research on the acceptance of telemedicine among older adults who are cancer patients.

Methods: We conducted a scoping review. PubMed, EMBASE, PsycINFO, CINAHL, and the Cochrane Central Register of Controlled Trials were systematically searched from inception to September 2020. Articles were included if the study population had a mean or median age ≥ 65 years, with cancer diagnoses and if the study assessed patients' acceptance of a telemedicine intervention. Quantitative, qualitative, and mixed method studies were included.

Results: Out of a total of 887 articles that were identified, 19 were included in the review. Interventions were delivered via telephone, videoconference, web portal, mobile app, wearable technology, and text messaging and included teleconsultation, monitoring and follow-up, psychosocial support and nursing care, and prompts. The most often cited facilitating factor was convenience. Other facilitators included an increase in telemedicine care accessibility, previous positive experiences of telemedicine, appropriate technical knowledge and support, decreased cost, physician recommendations, and privacy conferred by the telemedicine intervention. Barriers include a preference for conventional care along with negative perceptions of telemedicine, concerns about technical difficulties, and confidentiality concerns in the adoption of telemedicine.

Conclusions: None of the studies explored the ability of tailored interventions to address facilitators and barriers of the acceptance of telemedicine in order to increase its adoption by older adults. Facilitators and barriers will likely differ across different cultural contexts and by type of telemedicine; however, this is a gap in current knowledge. In-depth studies are necessary to determine if interventions could potentially address the barriers identified in this review, to increase acceptability.

(*J Med Internet Res* 2022;24(3):e28724) doi:[10.2196/28724](https://doi.org/10.2196/28724)

KEYWORDS

older adult patients; cancer; telemedicine; acceptability; satisfaction

Introduction

Cancer is among the most prevalent of noncommunicable diseases in high-income countries [1], and the general incidence is likely to rise with a greying population. Correspondingly, the number of older adults who are cancer patients requiring cancer care will only continue to increase [2]. That said, cancer care in the current era is increasing in complexity, with newer and better treatment options in aspects of care such as surgery, nutritional needs, nonsurgical treatment, wound care, and cancer surveillance [3]. Despite the amazing speed of technological adoption across different aspects of health care which are purportedly patient-centric (ie, bringing the service to the patients), care for patients with cancer remains very hospital-centric—the number of required visits to the hospital remains considerable for every patient. One mode of reducing hospital visits is the adoption of telemedicine, which has been accelerated by the COVID-19 pandemic [4-6].

Interestingly, there is no universal definition of *telemedicine*—one study [7] established 104 peer-reviewed definitions of the term. The World Health Organization defines *telemedicine* as “the delivery of health care services, where distance is a critical factor, by all health care professionals using information and communication technologies for the exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation, and for the continuing education of health care providers, all in the interests of advancing the health of individuals and their communities [8].” Based on this definition, telemedicine can take many forms, ranging from a simple phone call to televideo consultations and other potential applications.

While telemedicine could represent the next paradigm shift in global models of health care, barriers to its widespread adoption are present and can range from system-wide logistical issues to individual-level resistance, affecting patients, health care providers, and administrators. Crucially, while telemedicine adoption and use might be more intuitive in younger working adults who are likely to be more technologically savvy, it remains debatable how the incorporation of telemedicine into models of health care would be perceived among older adult patients, especially those with critical illness such as cancer [9].

The rapidly rising incidence of cancer among older adults [10] and the high demands of oncology care mean that this group of patients could potentially benefit immensely from the adoption of telemedicine; yet, there is also a risk that, instead, this population will be left behind by advances in technology. Telemedicine adoption is complex, and system- or community-level frameworks to explain socioenvironmental drivers—such as health economic or diffusion of innovation models—exist [11]. However, individual behavioral factors continue to play an important role in predicting telemedicine adoption at the inter- and intrapersonal level [11]. In this context, technology acceptance has often been identified as an important precursor to adoption of telemedicine initiatives. A comprehensive review [12] illustrated the applicability of technology acceptance models with respect to telemedicine acceptance; one of the most frequently utilized [12,13]

theoretical models identified was the Unified Theory of Acceptance and Use of Technology (UTAUT) [11]. While *acceptance* is not defined within the UTAUT model, the concept broadly involves an individual’s behavioral use of the technology or system [13].

In the UTAUT model, performance expectancy, effort expectancy, social influence, and facilitating conditions are 4 constructs that are direct determinants of behavioral intentions, and thus, of behavioral use: *performance expectancy* is defined as the degree to which an individual believes that using the technology will help in completing the task; *effort expectancy* is defined as the degree of ease of use of the technology; *social influence* is defined as the individual’s perception of whether important people around them believe that the individual should use the technology; and *facilitating conditions* is defined as the degree to which an individual believes that technical expertise and infrastructure are available to support the use of the technology [13].

Despite widespread use of *acceptance* as a concept to explain telemedicine adoption in the context of chronic diseases (eg, hypertension) and population (eg, caregivers, older adults), reviews, specifically on telemedicine for older adults who are patients with cancer [14-16], have not addressed telemedicine acceptance. Thus, we aimed to systematically map the international body of literature on acceptance of telemedicine among older adults who are patients with cancer and frame our current understanding of this topic within the UTAUT model.

Methods

Review Protocol and Search Strategy

Protocol

We based our review methodology and structure on an established example [17]. A protocol was developed *a priori*, and registered on PROSPERO (CRD42021235248); we elected to utilize PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-analysis for Scoping Reviews [17]) as the framework because the objective was to map existing evidence on telemedicine acceptance among older adult cancer patients, to identify the major knowledge gaps. The research question was defined using the PICOS (population, intervention, comparison, outcome, study design) framework. Specifically, we were interested in all quantitative, qualitative, or mixed methods studies (both observational and interventional studies) (study design) that examined older adult cancer patients’ (population) acceptance (outcome) of telemedicine initiatives (intervention). Where possible (eg, if a study utilized an interventional design), we were interested in telemedicine compared with standard (eg, face-to-face) care (comparison) (Table S1 in [Multimedia Appendix 1](#)).

Research Question

What is known from the literature about older adults (defined as ≥ 65 years old) who are cancer patients and their acceptance of the use of telemedicine?

Search Strategy

Based on the protocol, 2 reviewers (SYF, CYHW) conducted a comprehensive search of PubMed, EMBASE, PsycINFO, CINAHL, and the Cochrane Central Register of Controlled Trials. PubMed, EMBASE, and CINAHL were selected based on recommendations [18] on best databases for reviewing telemedicine-related topics. PsycINFO and the Cochrane Central Register of Controlled Trials were used because of the possibility that relevant studies examining telemedicine acceptance might have been published in behavioral psychology journals or as trials, respectively. Peer-reviewed papers were included if they were published before September 2020 and written in English. The concepts searched were relevant to telemedicine (eg, telemedicine, telehealth), attitudes toward telemedicine, and cancer patients (eg, neoplasm). Search terms were broad (for example, the Medical Subject Heading *attitude to health*) and applied to all fields (rather than simply to titles, abstracts, and article keywords) to minimize the likelihood of excluding relevant articles. The search strategy was constructed using free-text keywords and Boolean operators. The search strategy was refined through discussions among all coauthors (Table S2 in [Multimedia Appendix 1](#)).

To further reduce the likelihood of missing relevant articles, we manually searched the reference lists of articles included in the review.

Inclusion and Exclusion Criteria

First, as our study population of interest was older adults (defined as 65 years or older), we included only articles that had a study population with an average (mean or median) age that was 65 years or older. This ensured that there was a majority representation (at least 50% or more) of views of older adults in the reported outcomes. Studies that did not report mean or median age were, therefore, excluded. Second, only studies involving patients with cancer were included; cancer was defined as a malignant growth or tumor resulting from an uncontrolled division of cells. Third, studies must have included the use of telemedicine. Given that there has been no clear definition of telemedicine in the literature [7], in this review, any form of technology utilization to aid in delivery of health care was considered to be telemedicine. Fourth, patient acceptance of the telemedicine studied must have been reported. Acceptance was defined as an individual's likely behavioral use of the technology or system [13]. Participants may or may not have actually used the technology studied, as long as their perspective on the technology had been reported. Only studies reporting patient perspectives were included. Studies reporting

perspectives of health care providers or caregivers were excluded. Grey literature (eg, news articles, lecture slides, unpublished theses), reviews, editorials, letters to the editor, and studies not published in English or in peer-reviewed journals were excluded.

Selection of Articles

Duplicates were removed from the list of articles returned by the searches (manually and using EndNote X8). Two reviewers (SYF, CYHW) screened the titles and abstracts of all the articles independently to select articles for full-text review. Articles were included if at least one reviewer deemed it potentially suitable. Four reviewers (NQP, SYF, CYHW, JL) were involved in the review of full-text articles. Each full text was independently reviewed by at least 2 of the reviewers: full texts were evenly distributed among reviewers, and a second reviewer was assigned to independently verify each set, with moderate (based on [19]) levels of agreement overall (SYF and NQP: Cohen $\kappa=0.58$; CYHW and SYF: Cohen $\kappa=0.49$; JL and CYHW: Cohen $\kappa=0.46$; NQP and JL: Cohen $\kappa=0.86$). Disagreements were resolved by discussion, among all 4 reviewers, until consensus was reached.

Data Extraction and Analysis

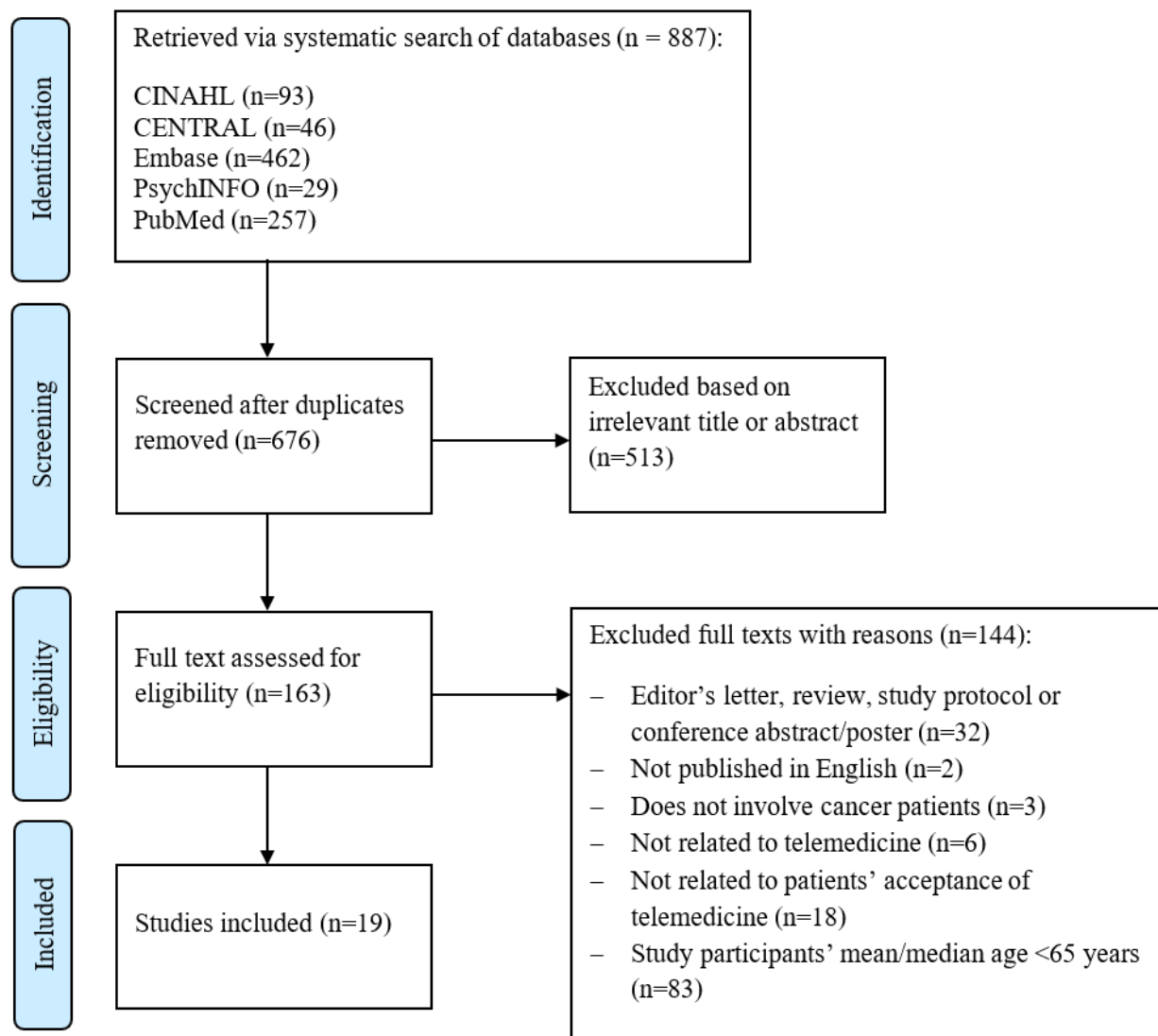
Data, relevant information on key study characteristics, and detailed information on the acceptance of telemedicine among older adults patients with cancer were extracted from the articles included in the review. Two reviewers (SYF, CYHW) independently charted data, discussed the results, and iteratively updated the data-charting form. The final data fields collected included the aim of the study, background or context, average (mean or median) age of participants, type of telehealth intervention (eg, teleconsultation, monitoring, nursing), telehealth technology used, presence of comparator arm, duration of intervention, outcome measured, methods of measuring outcomes (instrument or scales used), and key findings. Any disagreements were resolved by discussion or adjudication (by a third reviewer, JL).

Results

General

A total of 887 articles were identified, from which 211 duplicates were removed, leaving 676 articles for title and abstract screening. Of these, 513 articles were excluded, and 163 articles underwent full-text review, during which, 137 articles were excluded ([Figure 1](#); [Multimedia Appendix 1](#))

Figure 1. Study selection flowchart.



Descriptive Characteristics

More than two-thirds (14/19, 73.7%) of studies were performed in the United Kingdom and United States, and the rest were from other high-income countries; none were conducted in low- or middle-income nations. A clear increase in publication frequency was noted after the year 2010, with 21.1% (4/19) of studies published in the period from 2011 to 2015, and 57.9%

(11/19) of studies published in the period from 2016 to 2020. More than half (10/19, 52.6%) of the studies were cross-sectional (nonrandomized trial: 2/19, 10.5%; pre- and posttest: 1/19, 5.3%; randomized controlled trial: 2/19, 10.5%; qualitative: 4/19, 21.1%). Sample sizes ranged from n≤50 (6/19, 31.6%) to n>300 (1/19, 5.3%). The 19 studies included a range of different cancers (Table 1).

Table 1. Population and study characteristics.

Characteristics	Articles (n=19), n (%)
Country	
Canada	1 (5.3)
Denmark	1 (5.3)
Germany	2 (10.5)
The Netherlands	1 (5.3)
United Kingdom	7 (36.8)
United States	7 (36.8)
Publication year	
1995-2000	2 (10.5)
2001-2005	1 (5.3)
2006-2010	1 (5.3)
2011-2015	4 (21.1)
2016-2020	11 (57.9)
Study design	
Cross-sectional	10 (52.6)
Nonrandomized trial	2 (10.5)
Pre- and posttest	1 (5.3)
Randomized controlled trial	2 (10.5)
Qualitative	4 (21.1)
Sample size	
≤50	6 (31.6)
51-100	4 (21.1)
101-200	6 (31.6)
201-300	2 (10.5)
>300	1 (5.3)
Site of cancer	
Breast	1 (5.3)
Colorectal	1 (5.3)
Endometrial	3 (15.8)
Esophagogastric	1 (5.3)
Hematological	1 (5.3)
Lung	1 (5.3)
Prostate	2 (10.5)
Skin	1 (5.3)
Mixed	8 (42.1)

Range of Technology

Studies included interventions delivered via telephone (6/19, 31.6%), videoconference (4/19, 21.1%), web portal (2/19, 10.5%), mobile app (3/19, 15.8%), wearable technology (1/19, 5.3%), or text message (1/19, 5.3%). Of the 19 articles, 2 articles (10.5%) surveyed patients on technology in general (ie, without specifying a technology). Telemedicine interventions included

teleconsultation (2/19, 10.5%), monitoring and follow-up (7/19, 36.8%), psychosocial support and nursing care (5/19, 26.3%), and prompts (eg, to increase treatment adherence) (2/19, 10.5%). Of the 19 articles, 3 articles (15.8%) did not specify a particular telemedicine intervention, and participants in 5 studies (26.3%) did not actually use the technology in question—they were surveyed on their acceptance of that particular technology based on their impression of its usage (Table 2).

Table 2. Technology and intervention characteristics.

Characteristic	Articles (n=19), n (%)
Technology type	
Telephone	6 (31.6)
Mobile text	1 (5.3)
Videoconferencing	4 (21.1)
Web portal	2 (10.5)
Mobile app	3 (15.8)
Wearable technology	1 (5.3)
Technology in general	2 (10.5)
Type of telemedicine intervention	
Teleconsultation	2 (10.5)
Monitoring and follow-up	7 (36.8)
Psychosocial support and nursing care	5 (26.3)
Prompts	2 (10.5)
Not specified or general	3 (15.8)
Actual usage of technology studied	
Yes	14 (73.7)
No	5 (26.3)

Facilitators and Barriers to Technology Acceptance

Four studies [20-23] used previously validated questionnaires. None used an evaluation tool based on a theoretical behavioral

framework. Study topics were categorized into themes (Table 3).

Of the 19 articles, 15 articles reported either facilitators or barriers to explain their findings on acceptance; the other 4 articles did not [20,21,24-26,28-37] (Table 4).

Table 3. Key telemedicine acceptance findings.

Type of telemedicine and reference	Type of study	Aim of study	Key findings
Teleconsultation			
Allen et al (1995) [25]	Pre- and posttest study (n=21)	To assess levels of satisfaction (acceptance) among rural cancer patients being seen for clinic visits by using interactive videoconferencing	<ul style="list-style-type: none"> Patients were less inclined to want to use video system again when asked after attending an on-site consultation (P=.016)
Mair et al (2000) [31]	Qualitative (n=22)	To understand patients' views of telemedicine consultations	<ul style="list-style-type: none"> Participants felt that there was a difference in telemedicine consultations vs face-to-face visits (eg, modified behavior of patients) but were accepting of these differences to increase their access to health care and medical expertise
Monitoring and follow-up			
Overend et al (2008) [32]	Cross-sectional (n=53)	To determine whether nurse-led telephone clinic could effectively and safely be used to follow patients with indolent and chronic hematological malignancies	<ul style="list-style-type: none"> 62% of participants felt strongly that they would participate in a Teleclinic again rather than travel to see their oncologist
Beaver et al (2011) [20]	Cross-sectional (n=187)	To explore patient satisfaction on different aspects of follow-up service provision following treatment for colorectal cancer and amenability to an alternative strategy for follow-up care	<ul style="list-style-type: none"> 66% of patients mentioned that they would be willing to receive telephone follow-up care in the future Male patients were 2 times more likely to indicate willingness for telephone follow-up
Verma et al (2015) [27]	Cross-sectional (n=134)	To evaluate a radiographer-led telephone follow-up for patients with low to intermediate risk prostate cancer patients completing radiotherapy	<ul style="list-style-type: none"> 67 out of 88 (76%) expressed a preference for telephone follow-up, while 7% expressed no preference between clinic or telephone follow-up and 5.5% expressed preference for outpatient clinic follow-up
Onuma et al (2019) [38]	Cross-sectional (n=271)	To determine patient preferences around the preferred means of receiving information about cancer surveillance (secure digital communication versus phone call or office visit)	<ul style="list-style-type: none"> Patients >65 years preferred telephone or in-person communication of normal imaging results (ORa 2.03, 95% CI 1.16-3.56, P<.05) versus patients ≤65 years; all patients preferred telephone or in-person consult for abnormal results
Smits et al (2015) [22]	Nonrandomized trial (n=296)	To evaluate the effect of nurse led follow-up on quality of life and patient satisfaction compared to conventional follow-up in women treated for endometrial cancer and to evaluate the patient acceptance of nurse-led follow-up	<ul style="list-style-type: none"> Majority of women (98%) in the nurse-led follow-up group stated that they would like to continue their follow-up care with the nurse-led telephone clinic Women in both groups reported equal satisfaction with care
Wynter-Blyth et al (2017) [30]	Pre- and posttest (n=9)	To explore the perioperative potential of home remote-monitoring (eg, on adherence to prehabilitation and rehabilitation programs)	<ul style="list-style-type: none"> All 9 patients mentioned that they would recommend home remote-monitoring to other patients and 8 out of 9 said that they would consider buying their own personal health monitoring devices Helped to build their confidence in managing their condition and allowed them to play a more active role in their overall health, such as improving their adherence to exercise and diet
Nugteren et al (2017) [26]	Qualitative (n=20)	To investigate patients' opinions about the use of eHealth apps to support self-management in survivorship care	<ul style="list-style-type: none"> Majority of participants would like to use the app as they have positive attitudes toward the app
Psychosocial support and nursing care			

Type of telemedicine and reference	Type of study	Aim of study	Key findings
Xu et al (2014) [21]	Cross-sectional (n=230)	To model intention of lung cancer patients to using face-to-face and on-line lung cancer support groups	<ul style="list-style-type: none"> Positive intentions to join an online support group were reported by 34% of participants, whereas for the face-to-face support group, positive intentions to join were reported by 36.4% of participants
Beaver et al (2020) [34]	Cross-sectional (n=211)	To explore the preferences of endometrial cancer patients and their levels of satisfaction with hospital vs nurse-led telephone follow-up	<ul style="list-style-type: none"> Participants tended to prefer what was familiar to them; those in the hospital follow-up group tended to prefer hospital-based appointments while the telephone follow-up group tended to prefer appointments with a clinical nurse specialist, regardless of locality
Bohnenkamp et al (2004) [28]	Nonrandomized trial (n=28)	To measure the impact of telenursing on patients discharged with ostomies resulting from cancer treatment (telenursing + home health visit vs only home health visit)	<ul style="list-style-type: none"> 87% said they would prefer telenursing visit over waiting for a face-to-face visit 70% said that they would prefer a face-to-face visit (if no waiting time required) even though 85% agreed that the telenursing visit was as good as a face-to-face visit 93% of patients were satisfied with the telenursing combined with home health visit, while 81% were satisfied with just the home health visit ($P < .01$)
Bouchard et al (2019) [33]	Randomized controlled trial (n=192)	To examine the acceptability and efficacy for reducing disease-specific distress of a tablet-delivered psychosocial intervention for older men with advanced prostate cancer	<ul style="list-style-type: none"> Average exit survey responses were favorable and similar for intervention (mean 3.53, SD 0.55) and control (mean 3.65, SD 0.41; $P > .05$)
Williamson et al (2018) [35]	Qualitative (n=25)	To explore the views of women with endometrial cancer who had received telephone follow-up compared to hospital follow-up	<ul style="list-style-type: none"> Patients generally preferred telephone follow-up compared to hospital follow-up
Prompts			
Spoelstra et al (2016) [24]	Randomized controlled trial (n=75)	To conduct a preliminary evaluation of the efficacy of telemedicine with respect to adherence and symptom severity and interference in adult cancer patients prescribed Oral Anti-cancer medication	<ul style="list-style-type: none"> 97.4% recommended it as a way to assist patients to remember to take medications and 100% would recommend it to their oncologist as a way to monitor adherence 85.7% of participants completed the entire telemedicine intervention, suggesting that there is high acceptability Majority of participants (92.2%) reported high satisfaction
Brett et al (2018) [36]	Qualitative (n=18)	To assess the likely acceptability of an eHealth app in women who have utilized the app to support women prescribed adjuvant endocrine therapy after treatment for breast cancer	<ul style="list-style-type: none"> All participants except one said that they would recommend the app to women taking adjuvant endocrine therapy
Not specified or general			

Type of telemedicine and reference	Type of study	Aim of study	Key findings
Steeb et al (2019) [37]	Cross-sectional (n=200)	To investigate patient attitudes and their awareness toward skin cancer-related apps	<ul style="list-style-type: none"> • Most patients (86/126, 68.3%) rated scientifically reliable information as the most important feature for health-related apps, followed by user convenience (76/126, 60.3%) and data security (76/126, 60.3%) • For 54.0% (68/126) of patients, credibility of the app provider was important • 29.6% (37/125) and 25.4% (32/126) considered a low price and an attractive layout as critical, respectively
Rossen et al (2019) [23]	Cross-sectional (n=305)	To get insight of how receptive cancer survivors are toward using health technology for physical activity rehabilitation	<ul style="list-style-type: none"> • 88 participants (28.9%) were unreceptive toward supplementing their rehabilitation with technology devices
Rodler et al (2020) [29]	Cross-sectional (n=92)	To determine patients' perspective on adoption of telehealth as a response to the COVID-19 pandemic	<ul style="list-style-type: none"> • General sustainability of telehealth beyond pandemic: majority (65.9%) not inclined to continue telehealth measures • Type of treatment plays a role in telehealth acceptance: patients on immunotherapy are more willing to continue with telehealth measures than patients on chemotherapy

^aOR: odds ratio.

Table 4. Facilitators and barriers to telemedicine acceptance.

Reference	Facilitators	Barriers
Allen et al (1995) [25]	N/A ^a	<ul style="list-style-type: none"> Patients found that it was more difficult to be completely candid over the video consult than during the in-person consultation, when asked after the on-site consultation (P=.024)
Mair et al (2000) [31]	<ul style="list-style-type: none"> 100% of patients expressed positive attitudes regarding satisfaction with telemedicine; mainly due to convenience of access 	<ul style="list-style-type: none"> 50% expressed confidentiality concerns, and 50% felt telemedicine cannot fully replace face-to-face consults 41% were uneasy with the nurse as proxy for physical exam
Overend et al (2008) [32]	<ul style="list-style-type: none"> 78% of participants felt strongly that the teleclinic was convenient and/or saved them time and money 	<ul style="list-style-type: none"> Younger patients who lived one to two hours away from the cancer center declined participation in the teleclinic, as they did not consider the distance an inconvenience Some patients took the opportunity to shop when they came for follow-up visits and did not regard it always as an inconvenience
Beaver et al (2011) [20]	<ul style="list-style-type: none"> Greater satisfaction with the time given by professionals, practical advice, dietary information and comfort in contacting a colorectal nurse between appointments predicted for acceptance of telephone follow-up 	N/A
Verma et al (2015) [27]	<ul style="list-style-type: none"> 79 out of 88 patients (92%) who completed the satisfaction questionnaires reported that telephone follow-up was more or equally convenient as compared to clinic attendance 	N/A
Onuma et al (2019) [38]	N/A	N/A
Smits et al (2015) [22]	N/A	N/A
Wynter-Blyth et al (2017) [30]	<ul style="list-style-type: none"> Usability of the equipment was high, with 8 out of 9 participants finding the app and home remote-monitoring devices clear to navigate and easy to use 	<ul style="list-style-type: none"> 5 out of 9 patients experienced difficulties in reliability of the equipment, such as connection issues
Nugteren et al (2017) [26]	<ul style="list-style-type: none"> Perception that the app increases awareness of symptoms, concerns and supportive care possibilities and improves accessibility 	<ul style="list-style-type: none"> Concerns include: use of the app might be dependent on the current mood and state of the patient (eg, if unwell), preference for face-to-face contact, ease of use, need for concise and simple information, personalized advice as well as the importance of privacy
Xu et al (2014) [21]	Greater computer familiarity increases intention to join online support groups	<ul style="list-style-type: none"> Having more negative attitudes about online support groups decreases intention to join online support groups. Older age, being male and lower levels of education were associated with more negative attitudes about online support groups
Beaver et al (2020) [34]	<ul style="list-style-type: none"> telephone follow-up group more likely to indicate that follow-up appointments were always on time and more thorough compared to hospital follow-up group. Overall greater satisfaction with information received (more info is provided via telephone follow-up). Positive comments for telephone follow-up: knowing who to contact, convenience, being reassured 	<ul style="list-style-type: none"> Negative comments for telephone follow-up: preference for face-to-face contact, missing reassurance of clinical exam, feeling isolated/unsettled and problems with organizing telephone appointments
Bohnenkamp et al (2004) [28]	<ul style="list-style-type: none"> 100% of patients agreed that telenursing increased accessibility of care Less expenses due to less frequent pouch changes 	<ul style="list-style-type: none"> 15% of telenursing subjects reported that the camera and new technology embarrassed them
Bouchard et al (2019) [33]	N/A	N/A

Reference	Facilitators	Barriers
Williamson et al (2018) [35]	<ul style="list-style-type: none"> Telephone follow-up was more convenient for patients than hospital follow-up; patients did not have to rely on, or feel they were inconveniencing relatives or friends who would usually take them to hospital appointments, which promoted independence; easier to manage their day Patients found telephone follow-up reassuring and said they found it easier to self-manage than when they were receiving hospital follow-up telephone follow-up provided them with privacy that they perceived was not available at their hospital appointments 	N/A
Spoelstra et al (2016) [24]	N/A	N/A
Brett et al (2018) [36]	<ul style="list-style-type: none"> Phase 1: Women were generally positive about the concept of an app to provide info and support and all of them see great potential in the app for helping women cope with issues that may arise when taking adjuvant endocrine therapy; they also highlighted the accessibility of the app Phase 2: Women reported that the user interface was clear with intuitive controls and user satisfaction was good from the usability tasks Phase 3: Pilot of the app—the participants reported that downloading and navigating the app was straightforward and that it was user friendly 	Phase 1 device (eg, phone) required; app needed to be easy to download and simple to navigate.
Steeb et al (2019) [37]	<ul style="list-style-type: none"> 38.5% (72/187) thought that such apps are useful for patients; 42.6% (78/183) voted that skin cancer apps can supplement or support professional skin cancer screening by a physician 59.1% of the patients (110/186) would download a skin cancer app recommended by their physician Men were generally more willing to download an app that has been recommended by their physician than women (P=.02) 	<ul style="list-style-type: none"> 76.1% (140/184) figured that skin cancer apps cannot replace skin cancer screening performed by a physician Patients aged >61 years did not think that skin cancer apps can replace the physician in comparison to those under the age of 61 years (P=.02) and would rather read a printed brochure on skin cancer than download an app (P<.001)
Rossen et al (2019) [23]	<ul style="list-style-type: none"> Training and support in utilizing health technology for rehabilitation 	<ul style="list-style-type: none"> The unreceptive group has a higher representation of vulnerable individuals that are older, with lower educational level, live alone, currently smoke, and with more chronic comorbidities Unreceptive-group experience technology-specific barriers with significantly lower scores in dimensions related to their skills, motivation and user experiences
Rodler et al (2020) [29]	<ul style="list-style-type: none"> High response rates indicative of rapid acceptance of telehealth services by patients during pandemic despite difficulties of applicability in an aging population with lower email access or with hearing impairment; virtual communication was established quickly directly or through aiding relatives or partners 	<ul style="list-style-type: none"> Patients value personal interactions with their treating physicians greatly; patient–physician distancing can be perceived as a bigger toll than the risk of COVID-19

^aN/A: not applicable.

Facilitators

Five studies [27,31,32,34,35] found convenience to be a factor for technology acceptance, making it the most often cited facilitating factor. Four articles [26,28,29,36] mentioned increased accessibility to care as a reason for favoring telemedicine acceptance. Four articles [21,30,36,37] reported

that a positive experience of telemedicine increased the likelihood of technology acceptance, and 3 studies [21,30,36] stated the importance of having appropriate technical knowledge and support as a facilitating factor. Other factors included usability [30], lower cost [28], physician recommendation [37], and more privacy [35]. These facilitators are categorized within the theoretical framework of the UTAUT (Table 5).

Table 5. Facilitators and barriers within the context of the Unified Theory of Acceptance and Use of Technology model.

	Facilitators	Barriers
Performance expectancy	<ul style="list-style-type: none"> Increased accessibility Decreased cost Improved privacy 	<ul style="list-style-type: none"> Preference for conventional care Confidentiality concerns
Effort expectancy	<ul style="list-style-type: none"> Convenience Usability 	<ul style="list-style-type: none"> Technical difficulties
Social influence	<ul style="list-style-type: none"> Physician recommendation 	<ul style="list-style-type: none"> Negative perceptions of telemedicine
Facilitating conditions	<ul style="list-style-type: none"> Having appropriate technical know-how and support 	<ul style="list-style-type: none"> None found

Barriers

Six articles [21,25,26,31,34,37] found that preferences for conventional care along with negative perceptions of telemedicine were barriers to telemedicine acceptance. Participants from 4 articles [23,30,34,36] were concerned about technical difficulties, while those from another 4 articles [25,26,28,31] raised concerns about confidentiality in the adoption of telemedicine. These barriers are categorized within the theoretical framework of the UTAUT model (Table 5).

Acceptance Across Different Types of Telemedicine

In 2 early studies [25,31] on the acceptance of teleconsultation as a substitute for in-person consults, patients were hesitant to fully transition to teleconsultation.

Participant acceptance of technology in monitoring and follow-up of their oncological condition was explored in 7 studies [20,22,26,27,30,32,38]. Of these, 6 studies [18,20,22,27,32,33] reported a generally favorable attitude toward acceptance of telemedicine; none of these studies went as far as to conclude that telemedicine can be used to completely replace traditional in-person consult for follow-up care.

Acceptance of telemedicine related to psychosocial support and nursing care was studied in 5 studies [21,28,33-35]; 2 articles [28,35] reported a preference for telemedicine. The other 3 articles [21,33,34] reported no difference in preference between telemedicine versus control (traditional face-to-face support).

In both articles using technology as an adjunct for prompting adherence to treatment, participants were found to be generally receptive [24,36].

Three studies [23,29,37] explored participants' acceptance of technology without specifying the type of telemedicine intervention. One study [23] found that older patients were among those less likely to accept telemedicine as an adjunct to traditional care. Two studies [29,37] reported unfavorable acceptance of telemedicine as a replacement to conventional care. Of particular interest to the ongoing global pandemic was a study [37] on patients' perspectives on adoption of telemedicine as a response to the COVID-19 pandemic, in which it was noted that the majority of patients (65.9%) were not inclined to continue using telemedicine, because they greatly value personal interactions with their treating physicians. Steeb et al [37] concluded that patient-physician distancing can be perceived as eliciting a bigger toll than the risk of COVID-19.

Discussion

Principal Findings

To the best of our knowledge, this is the first scoping review to explore the acceptance of telemedicine among older adult patients with cancer. Our review of 19 primary studies indicated that numerous facilitators and barriers to acceptance of telemedicine exist. Because no study employed a theoretical behavioral framework in their methodology, it was not possible to map facilitators and barriers to UTAUT constructs with great precision. It was possible, however, to classify factors broadly according to the original definition [13] of each construct. We observed that all 4 UTAUT constructs play a role in determining behavioral use (Table 5).

Although facilitators and barriers to telemedicine acceptance in the older adult population were reported in most studies, there was a lack of studies employing methods beyond descriptive methods. None of the studies explored the effectiveness of tailored interventions to address facilitators and barriers to the acceptance of telemedicine by older adult populations; facilitators and barriers would likely differ across different cultural contexts and by type of telemedicine studied—this is a gap in our current knowledge of the use of telemedicine. Studies should aim to illustrate potential facilitators and barriers to telemedicine adoption while exploring the efficacy of different policies to address these factors within local settings. For example, common barriers identified in this review, such as negative perceptions of telemedicine, technical difficulties, and confidentiality concerns, can be easily targeted in education campaigns. The effectiveness of such campaigns can be evaluated and adapted to different local contexts. Policymakers may consider using the UTAUT model as a theoretical basis in designing such interventions.

While most studies showed positive acceptance of telemedicine as an adjunct to traditional care models, there was a paucity of articles that could confidently declare that the telemedicine alternative offered could completely replace the traditional modality. Older adult patients appear to still cherish the opportunity for face-to-face consults, and some studies [25,26,28,29,31,37] reported confidentiality concerns with technology. One article [37] was particularly illuminating—faced with the COVID-19 pandemic, a lot of older adult patients (who are also more vulnerable to complications of COVID-19) were forced to adopt telemedicine

as an alternative to in-person consultations, these older adult patients may otherwise not have tried this alternative. Nonetheless, the majority were not inclined to continue adopting such alternatives even if the pandemic continued [37]. This study [37] affords us an interesting view into older adult cancer patients' psyche in their attitudes toward telemedicine and suggests that more work is needed to examine the facilitators and barriers to telemedicine acceptance among older adults.

Five studies [22,24,28,31,34] reported level of satisfaction among their study population. Although there is generally high satisfaction reported in the use of telemedicine in these studies, acceptance of telemedicine among their study population remains inconclusive. For example, Mair et al [31] found that despite having all study participants reporting satisfaction with telemedicine care, 50% of participants felt telemedicine could not fully replace face-to-face consultations. While participants may be satisfied with the performance of a telemedicine technology, it does not naturally follow that they accept the use of telemedicine over conventional care methods. Acceptance of a telemedicine alternative and preference for its use over traditional care models are, therefore, different concepts from satisfaction. For telemedicine alternatives to enjoy widespread sustained adoption by the older adult patients when implemented by health care systems and organizations, future studies could examine the relationship between satisfaction and acceptance of telemedicine and analyze factors affecting the correlation between the two constructs.

The findings of our review may have implications for health care and public health policymakers in the context of telemedicine implementation. Caring for an older adult with cancer is complex. Health care innovation requires corresponding changes in behavioral patterns among the older adult population to realize its full potential. While some traditional barriers to health care access may be overcome, the acceptance of telemedicine itself is a behavior change, challenging the accepted health care delivery norm. In the adoption of telemedicine to aid in health care, the older adult population faces a whole new distinct set of facilitators and barriers, and an understanding of these is important to increase telemedicine acceptance [39]. As shown in our review, this may be done with the help of theoretical frameworks such as the UTAUT.

With the ongoing COVID-19 pandemic, which appears to be accelerating the worldwide adoption of telemedicine, it is inevitable that many patients will face the prospect of having to integrate telemedicine into their care routines—within the 6-month period from January to June 2020, there were 543 articles published on telemedicine-related literature [40]. New facilitators or barriers might emerge from this pandemic-induced adoption of telemedicine. Understanding these complex factors through the lens of a theoretical framework could provide a solid foundation for policymakers to navigate these challenging

times. Moving forward, it may also be worthwhile for public health researchers and professionals to adopt a mixed methods approach to understand how existing theories of technology acceptance may be extended to account for COVID-related factors. For example, the use of inductive grounded-theory methodologies may help to uncover novel themes and constructs not in existing theoretical frameworks.

Limitations

We acknowledge several limitations in our review. First, while we took the definition of older adults to be 65 years and older, it was not practical to identify only articles that contain a sample population of this age range. Doing so would have severely limited the number of articles available for analysis and would have excluded numerous studies in which the majority of the sample population were 65 years or older. We would have missed out on potentially available evidence in the literature pertaining to older adult patients and telemedicine. We, therefore, also included any studies in which a subgroup analysis on participants 65 years and older was performed. This compromise meant that a proportion (albeit, less than half) of the sampled population among included articles was younger than 65 years of age.

Second, the definition of telemedicine is very wide, and a range of telemedicine interventions had to be considered. This contributed to the heterogeneity of the studies included in this review.

Third, all the studies were conducted in Western high-income nations. This limits the generalizability of our findings. Attitudes of older adults with cancer toward telemedicine may be influenced by cultural factors and education level. The mean education levels of low- or middle-income nations could also be different from those of the sample populations included in this review; thus, the adoption of telemedicine in low- or middle-income nations by older adults with cancer might encounter additional difficulties.

Fourth, telemedicine is a rapidly emerging field, and its emergence has been expedited by the COVID-19 pandemic. While we were able to include articles from database inception to September 2020, it is likely that more studies have been published since September 2020.

Conclusion

The findings of our scoping review have important implications for future research. In a world grappling with the COVID-19 pandemic, telemedicine offers an alternative model of care and is here to stay. Our review has identified research gaps to be addressed. Future studies are necessary to understand the facilitators and barriers to telemedicine acceptance in older adults with cancer, with a view to investigating interventions to address barriers.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary tables.

[\[DOCX File, 18 KB - jmir_v24i3e28724_app1.docx\]](#)**References**

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2021 May;71(3):209-249 [[FREE Full text](#)] [doi: [10.3322/caac.21660](https://doi.org/10.3322/caac.21660)] [Medline: [33538338](https://pubmed.ncbi.nlm.nih.gov/33538338/)]
2. Takahashi K, Matsumoto H, Yamaguchi T, Nakano D, Watanabe F, Oohinata R, et al. [Surgery for colorectal cancer in elderly patients]. *Gan To Kagaku Ryoho* 2010 Dec;37(13):2823-2828. [Medline: [21160257](https://pubmed.ncbi.nlm.nih.gov/21160257/)]
3. Dodaro CA, Calogero A, Tammaro V, Pellegrino T, Lionetti R, Campanile S, et al. Colorectal cancer in the elderly patient: the role of neo-adjuvant therapy. *Open Med (Wars)* 2019 Aug;14:607-612 [[FREE Full text](#)] [doi: [10.1515/med-2019-0068](https://doi.org/10.1515/med-2019-0068)] [Medline: [31428685](https://pubmed.ncbi.nlm.nih.gov/31428685/)]
4. Hau YS, Kim JK, Hur J, Chang MC. How about actively using telemedicine during the COVID-19 pandemic? *J Med Syst* 2020 Apr 30;44(6):108 [[FREE Full text](#)] [doi: [10.1007/s10916-020-01580-z](https://doi.org/10.1007/s10916-020-01580-z)] [Medline: [32350626](https://pubmed.ncbi.nlm.nih.gov/32350626/)]
5. Shirke M, Shaikh S, Harky A. Implications of telemedicine in oncology during the COVID-19 pandemic. *Acta Biomed* 2020 Sep 07;91(3):e2020022 [[FREE Full text](#)] [doi: [10.23750/abm.v91i3.9849](https://doi.org/10.23750/abm.v91i3.9849)] [Medline: [32921719](https://pubmed.ncbi.nlm.nih.gov/32921719/)]
6. Boehm K, Ziewers S, Brandt MP, Sparwasser P, Haack M, Willems F, et al. Telemedicine online visits in urology during the COVID-19 pandemic?potential, risk factors, and patients' perspective. *Eur Urol* 2020 Jul;78(1):16-20 [[FREE Full text](#)] [doi: [10.1016/j.eururo.2020.04.055](https://doi.org/10.1016/j.eururo.2020.04.055)] [Medline: [32362498](https://pubmed.ncbi.nlm.nih.gov/32362498/)]
7. Sood S, Mbarika V, Jugoo S, Dookhy R, Doarn CR, Prakash N, et al. What is telemedicine? a collection of 104 peer-reviewed perspectives and theoretical underpinnings. *Telemed J E Health* 2007 Oct;13(5):573-590. [doi: [10.1089/tmj.2006.0073](https://doi.org/10.1089/tmj.2006.0073)] [Medline: [17999619](https://pubmed.ncbi.nlm.nih.gov/17999619/)]
8. Global health and ageing. World Health Organization. 2011. URL: https://www.who.int/ageing/publications/global_health.pdf [accessed 2021-11-20]
9. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. *J Telemed Telecare* 2018 Jan;24(1):4-12 [[FREE Full text](#)] [doi: [10.1177/1357633X16674087](https://doi.org/10.1177/1357633X16674087)] [Medline: [29320966](https://pubmed.ncbi.nlm.nih.gov/29320966/)]
10. Pilleron S, Sarfati D, Janssen-Heijnen M, Vignat J, Ferlay J, Bray F, et al. Global cancer incidence in older adults, 2012 and 2035: a population-based study. *Int J Cancer* 2019 Jan 01;144(1):49-58 [[FREE Full text](#)] [doi: [10.1002/ijc.31664](https://doi.org/10.1002/ijc.31664)] [Medline: [29978474](https://pubmed.ncbi.nlm.nih.gov/29978474/)]
11. Wade V, Gray L, Carati C. Theoretical frameworks in telemedicine research. *J Telemed Telecare* 2017 Jan;23(1):181-187. [doi: [10.1177/1357633X15626650](https://doi.org/10.1177/1357633X15626650)] [Medline: [26787643](https://pubmed.ncbi.nlm.nih.gov/26787643/)]
12. Venkatesh, Morris, Davis, Davis. User acceptance of information technology: toward a unified view. *MIS Q* 2003;27(3):425. [doi: [10.2307/30036540](https://doi.org/10.2307/30036540)]
13. 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](https://doi.org/10.2196/13117)] [Medline: [31115340](https://pubmed.ncbi.nlm.nih.gov/31115340/)]
14. Marzorati C, Renzi C, Russell-Edu SW, Pravettoni G. Telemedicine use among caregivers of cancer patients: systematic review. *J Med Internet Res* 2018 Jun 18;20(6):e223 [[FREE Full text](#)] [doi: [10.2196/jmir.9812](https://doi.org/10.2196/jmir.9812)] [Medline: [29914858](https://pubmed.ncbi.nlm.nih.gov/29914858/)]
15. Mileski M, Kruse CS, Catalani J, Haderer T. Adopting telemedicine for the self-management of hypertension: systematic review. *JMIR Med Inform* 2017 Oct 24;5(4):e41 [[FREE Full text](#)] [doi: [10.2196/medinform.6603](https://doi.org/10.2196/medinform.6603)] [Medline: [29066424](https://pubmed.ncbi.nlm.nih.gov/29066424/)]
16. 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](https://doi.org/10.2196/20359)] [Medline: [32784177](https://pubmed.ncbi.nlm.nih.gov/32784177/)]
17. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Tunçalp, 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](https://doi.org/10.7326/m18-0850)]
18. Bahaadinbeigy K, Yogesan K, Wootton R. MEDLINE versus EMBASE and CINAHL for telemedicine searches. *Telemed J E Health* 2010 Oct;16(8):916-919. [doi: [10.1089/tmj.2010.0046](https://doi.org/10.1089/tmj.2010.0046)] [Medline: [20925566](https://pubmed.ncbi.nlm.nih.gov/20925566/)]
19. Cohen J. A Coefficient of Agreement for Nominal Scales. *Educ Psychol Meas* 2016 Jul 02;20(1):37-46. [doi: [10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104)]
20. Beaver K, Wilson C, Procter D, Sheridan J, Towers G, Heath J, et al. Colorectal cancer follow-up: patient satisfaction and amenability to telephone after care. *Eur J Oncol Nurs* 2011 Feb;15(1):23-30. [doi: [10.1016/j.ejon.2010.05.006](https://doi.org/10.1016/j.ejon.2010.05.006)] [Medline: [20584628](https://pubmed.ncbi.nlm.nih.gov/20584628/)]
21. Xu Y, Testerman LS, Owen JE, Bantum EO, Thornton AA, Stanton AL. Modeling intention to participate in face-to-face and online lung cancer support groups. *Psychooncology* 2014 May 18;23(5):555-561. [doi: [10.1002/pon.3449](https://doi.org/10.1002/pon.3449)] [Medline: [24347369](https://pubmed.ncbi.nlm.nih.gov/24347369/)]
22. Smits A, Lopes A, Das N, Bekkers R, Kent E, McCullough Z, et al. Nurse-led telephone follow-up: improving options for women with endometrial cancer. *Cancer Nurs* 2015;38(3):232-238. [doi: [10.1097/NCC.000000000000177](https://doi.org/10.1097/NCC.000000000000177)] [Medline: [25098923](https://pubmed.ncbi.nlm.nih.gov/25098923/)]

23. Rossen S, Kayser L, Vibe-Petersen J, Ried-Larsen M, Christensen JF. Technology in exercise-based cancer rehabilitation: a cross-sectional study of receptiveness and readiness for e-Health utilization in Danish cancer rehabilitation. *Acta Oncologica* 2019 Jan 30;58(5):610-618. [doi: [10.1080/0284186x.2018.1562213](https://doi.org/10.1080/0284186x.2018.1562213)]
24. Spoelstra SL, Given CW, Sikorskii A, Coursaris CK, Majumder A, DeKoekkoek T, et al. Proof of concept of a mobile health short message service text message intervention that promotes adherence to oral anticancer agent medications: a randomized controlled trial. *Telemed J E Health* 2016 Jun;22(6):497-506 [FREE Full text] [doi: [10.1089/tmj.2015.0126](https://doi.org/10.1089/tmj.2015.0126)] [Medline: [26716365](https://pubmed.ncbi.nlm.nih.gov/26716365/)]
25. Allen A, Hayes J. Patient satisfaction with teleoncology: a pilot study. *Telemed J* 1995 Jan;1(1):41-46. [doi: [10.1089/tmj.1.1995.1.41](https://doi.org/10.1089/tmj.1.1995.1.41)] [Medline: [10165321](https://pubmed.ncbi.nlm.nih.gov/10165321/)]
26. Nugteren IC, Duineveld LAM, Wieldraaijer T, van Weert HCPM, Verdonck-de Leeuw IM, van Uden-Kraan CF, et al. Need for general practitioner involvement and eHealth in colon cancer survivorship care: patients' perspectives. *Fam Pract* 2017 Aug 01;34(4):473-478. [doi: [10.1093/fampra/cmw139](https://doi.org/10.1093/fampra/cmw139)] [Medline: [28122847](https://pubmed.ncbi.nlm.nih.gov/28122847/)]
27. Verma R, Treasure P, Hughes R. Development and evaluation of radiographer led telephone follow up following radical radiotherapy to the prostate. a report of a Macmillan Cancer Support Sponsored Pilot project at Mount Vernon Hospital. *Radiography* 2015 Feb;21(1):16-24. [doi: [10.1016/j.radi.2014.05.003](https://doi.org/10.1016/j.radi.2014.05.003)]
28. Bohnenkamp SK, McDonald P, Lopez AM, Krupinski E, Blackett A. Traditional versus telenursing outpatient management of patients with cancer with new ostomies. *Oncol Nurs Forum* 2004 Sep;31(5):1005-1010. [doi: [10.1188/04.ONF.1005-1010](https://doi.org/10.1188/04.ONF.1005-1010)] [Medline: [15378102](https://pubmed.ncbi.nlm.nih.gov/15378102/)]
29. Rodler S, Apfelbeck M, Schulz GB, Ivanova T, Buchner A, Staehler M, et al. Telehealth in uro-oncology beyond the pandemic: toll or lifesaver. *Eur Urol Focus* 2020 Sep 15;6(5):1097-1103 [FREE Full text] [doi: [10.1016/j.euf.2020.05.010](https://doi.org/10.1016/j.euf.2020.05.010)] [Medline: [32534969](https://pubmed.ncbi.nlm.nih.gov/32534969/)]
30. Wynter-Blyth V, MacKinnon T. Streamlining perioperative care for oesophago-gastric cancer surgery patients using home remote monitoring. *Prim Health Care* 2017 May 30;27(5):27-31. [doi: [10.7748/phc.2017.e1225](https://doi.org/10.7748/phc.2017.e1225)]
31. Mair F, Whitten P, May C, Doolittle GC. Patients' perceptions of a telemedicine specialty clinic. *J Telemed Telecare* 2000 Jun 24;6(1):36-40. [doi: [10.1258/1357633001933925](https://doi.org/10.1258/1357633001933925)] [Medline: [10824389](https://pubmed.ncbi.nlm.nih.gov/10824389/)]
32. Overend A, Khoo K, Delorme M, Krause V, Avanesian A, Saltman D. Evaluation of a nurse-led telephone follow-up clinic for patients with indolent and chronic hematological malignancies: a pilot study. *Can Oncol Nurs J* 2008;18(2):64-73. [doi: [10.5737/1181912x1826468](https://doi.org/10.5737/1181912x1826468)] [Medline: [18649698](https://pubmed.ncbi.nlm.nih.gov/18649698/)]
33. Bouchard LC, Yanez B, Dahn JR, Flury SC, Perry KT, Mohr DC, et al. Brief report of a tablet-delivered psychosocial intervention for men with advanced prostate cancer: acceptability and efficacy by race. *Transl Behav Med* 2019 Jul 16;9(4):629-637 [FREE Full text] [doi: [10.1093/tbm/iby089](https://doi.org/10.1093/tbm/iby089)] [Medline: [30285186](https://pubmed.ncbi.nlm.nih.gov/30285186/)]
34. Beaver K, Williamson S, Sutton CJ, Gardner A, Martin-Hirsch P. Endometrial cancer patients' preferences for follow-up after treatment: a cross-sectional survey. *Eur J Oncol Nurs* 2020 Apr;45:101722. [doi: [10.1016/j.ejon.2020.101722](https://doi.org/10.1016/j.ejon.2020.101722)] [Medline: [32014709](https://pubmed.ncbi.nlm.nih.gov/32014709/)]
35. Williamson S, Beaver K, Gardner A, Martin-Hirsch P. Telephone follow-up after treatment for endometrial cancer: a qualitative study of patients' and clinical nurse specialists' experiences in the ENDCAT trial. *Eur J Oncol Nurs* 2018 Jun;34:61-67. [doi: [10.1016/j.ejon.2018.02.005](https://doi.org/10.1016/j.ejon.2018.02.005)] [Medline: [29784140](https://pubmed.ncbi.nlm.nih.gov/29784140/)]
36. Brett J, Boulton M, Watson E. Development of an e-health app to support women prescribed adjuvant endocrine therapy after treatment for breast cancer. *Patient Prefer Adherence* 2018 Dec;Volume 12:2639-2647. [doi: [10.2147/ppa.s187692](https://doi.org/10.2147/ppa.s187692)]
37. Steeb T, Wessely A, Mastnik S, Brinker TJ, French LE, Niesert A, et al. Patient attitudes and their awareness towards skin cancer-related apps: cross-sectional survey. *JMIR Mhealth Uhealth* 2019 Jul 02;7(7):e13844 [FREE Full text] [doi: [10.2196/13844](https://doi.org/10.2196/13844)] [Medline: [31267978](https://pubmed.ncbi.nlm.nih.gov/31267978/)]
38. Onuma AE, Palmer Kelly E, Chakedis J, Paredes AZ, Tsilimigras DI, Wiemann B, et al. Patient preferences on the use of technology in cancer surveillance after curative surgery: a cross-sectional analysis. *Surgery* 2019 Apr;165(4):782-788 [FREE Full text] [doi: [10.1016/j.surg.2018.12.021](https://doi.org/10.1016/j.surg.2018.12.021)] [Medline: [30770135](https://pubmed.ncbi.nlm.nih.gov/30770135/)]
39. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. *Int J Med Inform* 2017 May;101:75-84. [doi: [10.1016/j.ijmedinf.2017.02.002](https://doi.org/10.1016/j.ijmedinf.2017.02.002)] [Medline: [28347450](https://pubmed.ncbi.nlm.nih.gov/28347450/)]
40. Doraiswamy S, Abraham A, Mantani 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](https://doi.org/10.2196/24087)] [Medline: [33147166](https://pubmed.ncbi.nlm.nih.gov/33147166/)]

Abbreviations

PICOS: population, intervention, comparison, outcomes, study design

UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Mavragani; submitted 11.03.21; peer-reviewed by S Bidmon, F Lei; comments to author 08.06.21; revised version received 11.10.21; accepted 19.02.22; published 29.03.22.

Please cite as:

Pang NQ, Lau J, Fong SY, Wong CYH, Tan KK

Telemedicine Acceptance Among Older Adult Patients With Cancer: Scoping Review

J Med Internet Res 2022;24(3):e28724

URL: <https://www.jmir.org/2022/3/e28724>

doi: [10.2196/28724](https://doi.org/10.2196/28724)

PMID: [35348462](https://pubmed.ncbi.nlm.nih.gov/35348462/)

©Ning-Qi Pang, Jerrald Lau, Si-Ying Fong, Celine Yu-Hui Wong, Ker-Kan Tan. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 29.03.2022. 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.

Review

Telehealth Business Models and Their Components: Systematic Review

Farnia Velayati¹, PhD; Haleh Ayatollahi², PhD; Morteza Hemmat³, PhD; Reza Dehghan⁴, PhD

¹Department of Health Information Management, School of Health Management and Information Sciences, Iran University of Medical Sciences, Tehran, Iran

²Health Management and Economics Research Center, Health Management Research Institute, Iran University of Medical Sciences, Tehran, Iran

³Department of Health Information Technology, Saveh University of Medical Sciences, Saveh, Iran

⁴Department of Health Entrepreneurship Management, Virtual University of Medical Sciences, Tehran, Iran

Corresponding Author:

Haleh Ayatollahi, PhD

Health Management and Economics Research Center

Health Management Research Institute

Iran University of Medical Sciences

No. 4 Rashid Yasemi St, Vali-Asr St

Tehran, 1996713883

Iran

Phone: 98 21 88794302

Fax: 98 21 88883334

Email: ayatollahi.h@iums.ac.ir

Abstract

Background: Telehealth technology is an excellent solution to resolve the problems of health care delivery. However, this technology may fail during large-scale implementation. As a result, business models can be used to facilitate commercialization of telehealth products and services.

Objective: The purpose of this study was to review different types of business models or frameworks and their components used in the telehealth industry.

Methods: This was a systematic review conducted in 2020. The databases used for searching related articles included Ovid, PubMed, Scopus, Web of Science, Emerald, and ProQuest. Google Scholar was also searched. These databases and Google Scholar were searched until the end of January 2020 and duplicate references were removed. Finally, articles meeting the inclusion criteria were selected and the Critical Appraisal Skills Programme (CASP) checklist was used for appraising the strengths and limitations of each study. Data were extracted using a data extraction form, and the results were synthesized narratively.

Results: Initially, 4998 articles were found and after screening, 23 were selected to be included in the study. The results showed that new telehealth business models were presented in 13 studies, and the applications of the existing business models were reported in 10 studies. These studies were related to different types of services, namely, telemonitoring (4 studies), telemedicine (3 studies), mobile health (3 studies), telerehabilitation (3 studies), telehealth (2 studies), assisted living technologies (2 studies), sensor-based systems (2 studies), and mobile teleradiology, teleradiology, telecardiology, and teletreatment (1 study related to each area). In most of the business models, value proposition, financial variables, and revenue streams were the main components.

Conclusions: Applying business models in the commercialization of telehealth services will be useful to gain a better understanding of the required components, market challenges, and possible future changes. The results showed that different business models can be used for different telehealth technologies in various health systems and cultures. However, it is necessary to evaluate the effectiveness of these models in practice. Moreover, comparing the usefulness of these models in different domains of telehealth services will help identify the strengths and weaknesses of these models for future optimization.

(*J Med Internet Res* 2022;24(3):e33128) doi:[10.2196/33128](https://doi.org/10.2196/33128)

KEYWORDS

telehealth; telemedicine; mobile health; business model; value; commerce; revenue; market; systematic review; health care

Introduction

Currently, health care systems are experiencing significant increases in costs mainly due to the shortage of health care professionals, increasing life expectancy, growing elderly population, and identification of new diseases and treatment methods [1,2]. In addition, economic developments, improved quality of life, and better health conditions along with more efficient policy making have led to a demographic transition (ie, an increase in the elderly population and a reduction in the young population) [3]. To resolve the challenges associated with health care delivery to different groups of patients, information technology-based solutions such as telehealth technology have been proposed [4-6]. Telehealth is defined as the use of information and communication technology to provide a wide range of health care services [7,8]. Telehealth has also been considered a unique opportunity to bridge the gaps and inequalities in health care delivery and as a solution to reduce the pressure imposed on health care systems [9,10]. It should be noted that the term telehealth includes telemedicine, eHealth, and mobile health (mHealth), and these terms are sometimes used interchangeably [1,5,11,12].

Currently, commercialization in the telehealth industry has received significant attention and innovative technology-based start-ups are expanding. In fact, the real value of these innovations lies in their commercialization [13-17]. The results of various studies show that the use of innovative technologies in the fields of telehealth and telemedicine is very challenging, and many products in these fields either fail in the implementation phase or stop in the research and development phase [13-16]. Most of these innovations and new technologies have never been introduced at the market level, as they have mainly focused on technology-based solutions rather than real value [1,14,15]. There are also a number of nontechnical challenges such as the nature of the relationship between health care providers and patients, the responsibility of information technology professionals, and privacy and confidentiality issues that should not be underestimated [18]. To overcome these challenges, the use of business models seems inevitable for successful commercialization of innovative technologies, and it may lead to more effective and efficient provision of health care services [17].

Recently, different business models have been proposed for the telehealth industry [1,19]. However, the findings of the research conducted by Frederickson et al showed that a business model and its components should be chosen based on the purpose of the technology and the context of use [20]. The results of other studies have indicated that patients, health care providers, payers, vendors, and other stakeholders play a key role in providing telehealth and telemedicine services. If a business model provides social or economic value for all stakeholders, then the likelihood of the successful implementation of a technology will increase [13,21]. It should be noted that different business models may have different components, as reported in various studies [22,23]. As successful commercial investment in telehealth requires an appropriate business model and plan, understanding these models and their components will help technology developers and commercial investors to make more

informed decisions in this field [22,23]. Therefore, the purpose of this study was to review different types of business models and their components used in the telehealth industry.

Methods

This study was a systematic review conducted in 2020. A systematic review is a type of review in which a systematic method is used to summarize evidence on questions with a detailed and comprehensive method [24]. Before conducting this review, ethical approval was obtained from the ethics committee of the Iran University of Medical Sciences (reference number: IR.IUMS.REC.1397.1328).

Eligibility Criteria

To select the most relevant studies, inclusion and exclusion criteria were set. According to the inclusion criteria, all research papers, reviews, conference papers, theses, and dissertations in which business models or business frameworks and their components were discussed in relation to telehealth, telemedicine, and mHealth were included in the study. No time frame was considered for searching the articles and the search was conducted until the end of January 2020. Papers published in English and full-text availability were the other inclusion criteria.

According to the exclusion criteria, books, book chapters, letters to the editor, and studies in which a business model or framework was used in fields other than telehealth, telemedicine, and mHealth were excluded. Publication languages other than English and unavailability of full texts were the other exclusion criteria.

Information Sources

The databases used for searching articles included Ovid, PubMed, Scopus, Web of Science, Emerald, and ProQuest. Google Scholar was also searched. The searches were conducted until the end of January 2020 and duplicate references were removed. Additionally, the OpenGrey database was searched to find grey literature. The search process was carried out by reference and citation tracking, and the scientific profiles of the authors of the articles were examined to find further related articles.

Search Strategy

To develop a search strategy, MeSH (Medical Subjects Headings) terms such as commerce, mHealth, and telemedicine and key terms such as business, business model, value chain, eHealth service, and commercial phenomena were used. The search strategies, number of records, and search dates are presented in [Multimedia Appendix 1](#). There was no time limit for searching the articles, but the language was limited to English and only full-text papers were included in the study.

Selection Process

The screening process was performed based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) checklist [25,26]. After retrieving relevant articles, reference management was performed using EndNote software (Version X7, Clarivate) and duplicates were removed.

The titles, abstracts, and full texts of the retrieved studies were screened. The initial search and screening processes were conducted by one of the authors (FV). In the next step, the other authors independently screened and appraised the remaining articles and resolved discrepancies by discussion and reached a consensus.

Data Collection Process

Data were extracted using a data extraction form comprising the name(s) of the author(s), year of publication, country, research objective, research method, business model, the model's components, and a summary of the results. The first author (FV) initially collected the data, and the reports were reviewed independently by other researchers too. In case of disagreement, the researchers discussed the issue and resolved it by reaching a consensus.

Data Items

In this study, the business models or frameworks and their components used for the commercialization of telehealth services were the main data items that were reviewed and compared in different studies.

Risk of Bias Assessment

The quality of the studies was assessed using the Critical Appraisal Skills Programme (CASP) checklist [27]. As qualitative methods were used in the selected studies, the CASP checklist for qualitative research was used. It consists of 10 questions, with "yes," "no," or "can't tell" as the answer options. The calculated scores showed if the quality of each study as high (7-10), medium (4-6), or low (1-3). The assessment was

conducted by 2 researchers (FV and HA) independently (see [Multimedia Appendix 2](#)).

Synthesis Methods

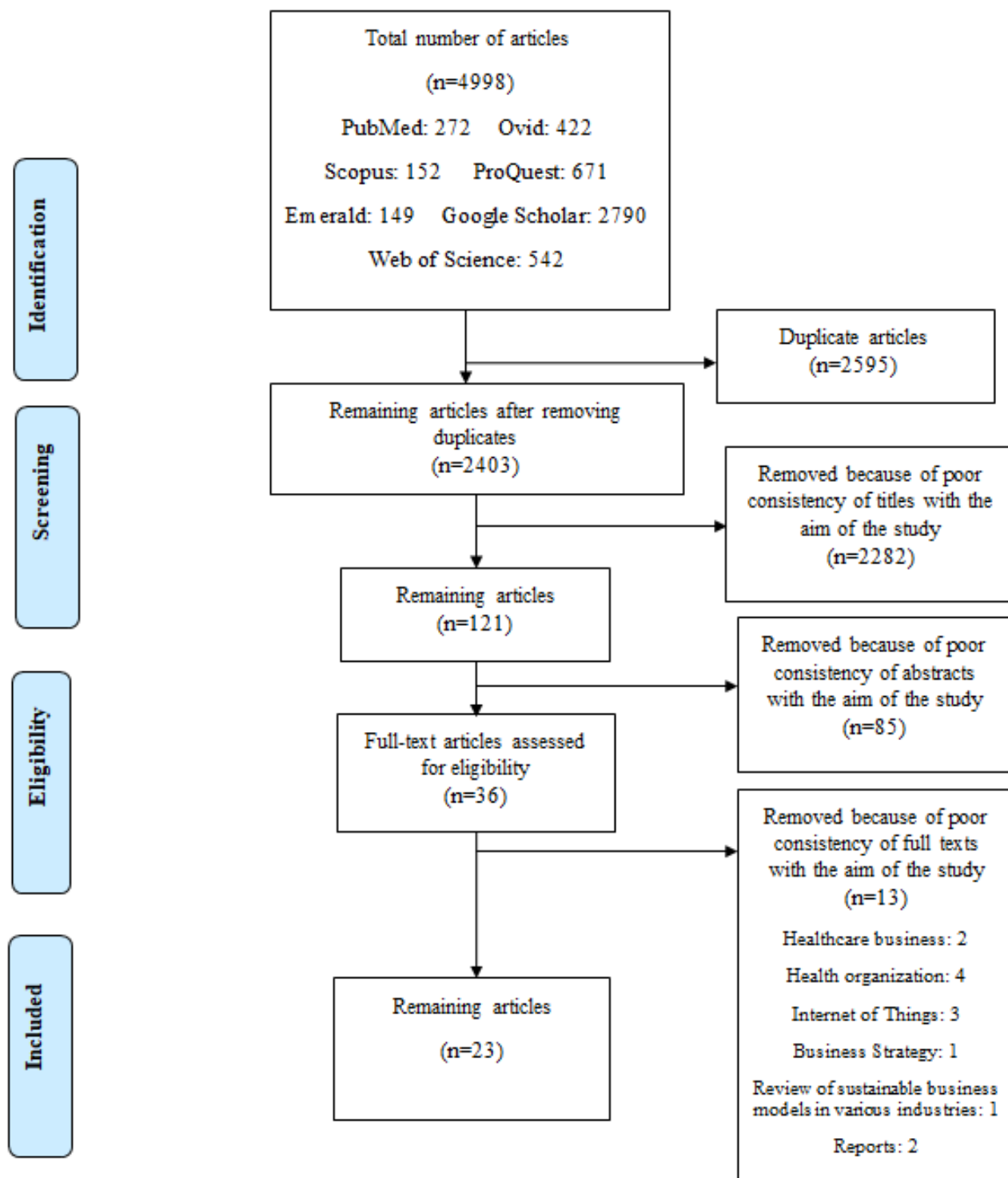
In most of the selected studies, qualitative methods were used. Therefore, meta-analysis was not possible. The papers were divided into 2 groups. The first group included those studies in which a new business model or framework was presented, and the second group included papers analyzing existing business models used in the telehealth industry. To summarize data, tables were developed based on the data extraction form. The main components of the business models are also presented in figures for better understanding.

Results

Study Selection

The preliminary search results in the selected databases included 4998 articles. After excluding duplicates, 2403 articles remained. Then, the titles of these articles were reviewed, and 2282 articles were excluded due to poor alignment of their aims with those of this study. In the next step, the abstracts of the 121 remaining articles were reviewed and 85 papers were excluded because their content was mostly irrelevant to the aims of this study. The full texts of the remaining articles (n=36) were reviewed, and 13 articles were excluded as they were mainly related to health care businesses (n=2), health organizations (n=4), Internet of Things (n=3), business strategy (n=1), sustainable business models in various industries (n=1), and organizational reports (n=2). Finally, 23 papers were selected to be included in this review. [Figure 1](#) shows the article selection process.

Figure 1. Article selection process based on PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [25-26].



Study Characteristics

The papers selected for inclusion in this review were divided into 2 main groups. The first group included 13 studies that presented a new business model or framework [1,14,19,23,28-36], and the second group comprised 10 studies that evaluated existing business models or frameworks used in the telehealth industry [37-46]. These studies were published between 2005 and 2020. The first group comprised studies conducted in the United States (n=5), the Netherlands (n=1),

Germany (n=2), Taiwan (n=3), Italy (n=1), and England (n=1); the second group included studies conducted in the Netherlands (n=6), China (n=1), Australia (n=2), and Sweden (n=1).

Results of Individual Studies

New Business Models or Frameworks for Telehealth Industry

Table 1 presents the 13 studies analyzing a new business model or framework for use in the telehealth industry [1,14,19,23,28-36].

Table 1. Summary of the studies presenting a new business model or framework for use in the telehealth industry.

Author(s) (year) and country	Objective	Methods	Business model	Model components	Results
Barker et al (2005) [23] United States	To describe a business model that was developed specifically to distribute telemedicine services throughout the state of Arizona at the lowest possible cost	Qualitative study (case study)	Arizona Telemedicine Program (ATP) business model	Five components: vendor services layer, infrastructure services layer, operational services layer, professional services layer, and client layer	The ATP business model was a layered model where each layer supported the upper layer, and the membership model has allowed the ATP to develop a modern telecommunication network that delivers services to clients at a lower cost because of its distributed network and services.
Mun et al (2005) [28] United States	To provide a business model of teleradiology	Qualitative study (literature review)	Teleradiology business model	Five teleradiology business models: stand-alone teleradiology practice, the “nighthawk”/on-call coverage, solo radiologist practice, expert/second-opinion teleradiology, and a global virtual radiology service based on workload sharing and reallocation	This successful business model will depend on the ability to produce the highest-quality product at the lowest cost.
Fife and Pereira (2008) [30] United States	To provide a VISOR ^a business model framework for mobile telehealth	Qualitative study (case study)	VISOR business model framework	Five components: value proposition, interface, service platforms, organizing model, and revenue model	The VISOR framework suggests that widespread adoption of mobile health care can only be achieved when the interface, service platform, organizational model, and revenue model are addressed simultaneously.
Kijl et al (2010) [14] Netherlands	To design a business model for a myofeedback-based teletreatment service (MyoTel) in patients with chronic neck and shoulder pain or whiplash injury	Mixed methods study (quantitative and qualitative case study)	Abstract cost benefit model (ACBM)	Two components: demand and supply	The business model engineering strategy provided important insights that led to an improved, a more viable, and a feasible business model; the related value network design and the process of early-stage business model engineering reduce risk and produce substantial savings in costs and resources related to service deployment.
Lin et al (2010) [32] Taiwan	To analyze the business model of a service innovation case by evaluating a telecardiology service	Qualitative study (case study)	Telecardiology business model	Nine components: value proposition, target customer, distribution channel, (customer) relationship, value configuration, capability, partnership, cost structure, and revenue model	The telecardiology service continued to succeed because of the mutual benefits it offered to the providers and users. A telecare service is meaningful to the general public only when the business model is sustainable.
Lin et al (2011) [33] Taiwan	To generate a framework to analyze 6 major telemedicine projects in Taiwan	Qualitative study (interviews with hospitals, security firms, and not-for-profit organizations)	Telemedicine framework	Six components: value proposition, target customers, service process, resources and capabilities, partnership and cost structure, and revenue model	Value proposition, partnership, resource, and capability affect service processes and cost structures. This in turn impacts customers' acceptance of telemedicine.

Author(s) (year) and country	Objective	Methods	Business model	Model components	Results
Fachinger and Schöpke (2014) [34] Germany	To identify, describe, and develop business models of sensor-based fall detection systems	Qualitative study (literature review)	Consistent business model	Nine partial models: customer, market, financing, proceeds, production, resources, procurement, network, and strategy	A sustainable business model was built by interconnecting 9 partial business models.
Peters et al (2015) [19] Germany	To describe, analyze, and classify a business model	Qualitative study (literature review)	CompBizMod framework	Four components: value proposition, value co-creation, value communication and transfer, and value capture	This business framework was useful for coordinating the perspectives of different telemedicine institutions, evaluating competitors, and designing competitive advantages.
Fusco and Turchetti (2015) [31] Italy	To identify the best business model to optimize value creation for most project stakeholders	Qualitative study (interviews with decision makers, physiotherapists, patients, and caregivers)	Telerehabilitation business/governance models	Three components: key activities, customer/patient segments, and key resources	Telerehabilitation business models reduced costs and the number of people on the waiting list. Actually, due to changes in the health sector and innovative governance, patients can be involved in the recovery process.
Lee and Chang (2016) [29] Taiwan	To find a business model to improve the health management of patients with chronic kidney disease	Qualitative study (literature review)	Mobile health management business model	Four components: data, data analysis/service, user, and partner	Requirement analysis and design of the mobile health management business model led to the provision of a cheap and professional support and management services platform for the disease.
Oderanti and Li (2016) [35] England	To investigate the commercialization of assisted living technologies and provide a sustainable business model	Qualitative study (literature review)	Sustainable business model	Seven components: value proposition, product innovation and commercialization, infrastructure management, customer relations management, financial viability and sustainability, stakeholder credibility, and revenue streams	The comparative advantage of a sustainable business framework was the most important factor that encouraged older people to pay for eHealth despite their free health services. Further, this sustainable model reduced the pressure on the British health system.
Pereira (2017) [1] United States	To identify the value proposition of telehealth	Qualitative study (literature review)	VISOR business model	Five components: value proposition, interface, service platforms, organizing model, and revenue model	The VISOR framework illustrates that although technology issues, such as security and privacy considerations, remain key factors that will determine the rate of adoption of telehealth, non-technological challenges are equally, if not more, important.
Arkwright et al (2019) [36] United States	To provide a business model for the success of telehealth programs	Qualitative study (literature review)	Telehealth business model	Eight models: direct-to-consumer (patient), organization-to-organization, clinician-to-clinician, oversight and processes, online patient access/portals/technology, mHealth/medical applications, hardware/software, and international telehealth program	A successful telemedicine business model must be safe, appropriate for the patient's needs, patient-centered, user-friendly, compliant, mission driven/strategically aligned, and have demonstrable value for the patient.

^aVISOR: value proposition, interface, service platform, organizing model, and revenue.

The study by Barker et al [23] presents a 5-layer model for telemedicine. From the bottom to the top, these layers included the vendor services layer, infrastructure services layer, operational services layer, professional services layer, and client layer. In this model, each layer supported the top layer, and the model created a new and low-cost infrastructure for telecommunication by developing a membership program and connecting to other networks. It also led to the distribution of specialized clinical services in rural communities [23]. Mun et al presented 5 business models for teleradiology including stand-alone teleradiology practice, the “nighthawk” or on-call coverage, solo radiologist practice, expert or second-opinion teleradiology, and a global virtual radiology service based on workload sharing and reallocation. These models led to more effective, higher-quality, and less-expensive diagnoses [28].

In 2 studies, business models were presented for mHealth services [29,30]. Among these, the study conducted by Fife and Pereira used the 5-component VISOR (value proposition, interface, service platform, organizing model, and revenue) model as the analytical framework to identify and address barriers to the widespread use of telehealth [30]. Another study was conducted by Lee and Chang that provided designing a 4-component business model for mHealth services for chronic kidney disease. The 4 components of this model were data, data analysis/service, partner, and user, which finally provided a cost-effective and professional platform for disease support and management services [29].

In the field of telerehabilitation, 2 different business models were presented in 2 of the included studies [14,31]. In the study by Kijl et al, a business model was considered for treating patients with chronic pain in the shoulder and neck. The design of this business model included a demand component on one side and a supply component on the other side. Medical research and development organization, occupational health care organization, and disability insurance organization were the subsets of supply and demand. In the value network of this business model, the components were interrelated, and increased productivity compensated for the additional costs of information technology [14].

Fusco and Turchetti also presented 4 models of business governance for telerehabilitation after total knee replacement. These models included 1 conservative model, 2 partnership models between primary care units and supporting companies that supplied equipment for primary care units, and 1 model based on cooperation between stakeholders. The results showed that the innovation structure was enhanced from the first to the fourth business model. The main components of these models were key activities, customer and patient segments, and key resources. These models reduced costs and the number of people on the waiting list [31].

In telecardiology, the results of the study by Lin et al showed that using a sustainable business model with the 9 components including value proposition, target customer, distribution channel, (customer) relationship, value configuration, capability,

partnership, cost structure, and revenue influenced the acceptance of technology by the general public and provided mutual benefits for service providers and patients [32].

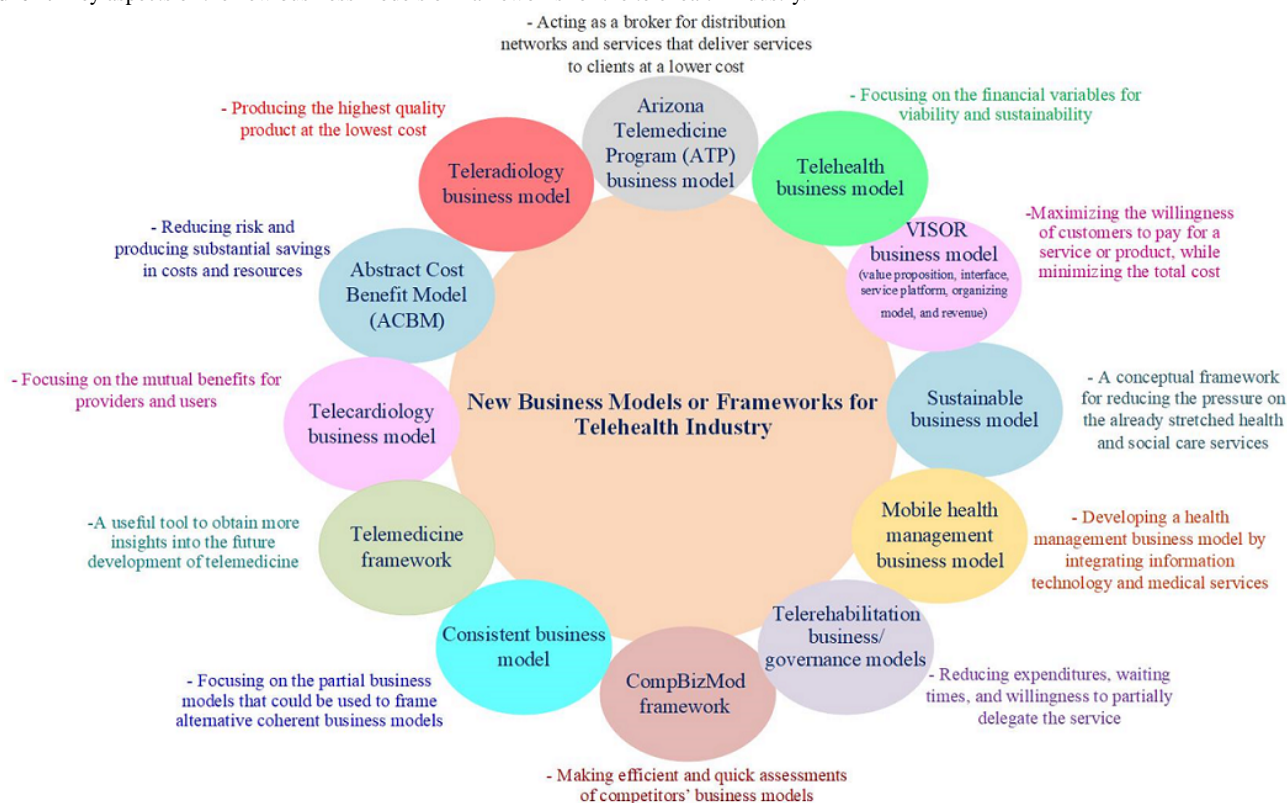
In 2 studies, a business framework for telemedicine was presented [19,33]. In the study of Lin et al, the business framework included the components of value proposition, partnership, resources and capabilities, and geography [33]. In another study, Peters et al revealed that the CompBizMod framework in telemedicine created a new perspective for reviewing and evaluating current business models in terms of structure, logic, and value. This framework included 4 main components of value proposition, value co-creation, value communication and transfer, and value capture, and the framework could be used to generate different perspectives in telemedicine business models, evaluate competitors, and determine competitive advantages [19].

The results of the study conducted by Fachinger and Schöpke showed that a sustainable business model in sensor-based fall recognition systems consists of 9 interconnected components, building blocks, or partial models including customer, market, financing, proceeds, production, resources, procurement, network, and strategy; the combined application of these components led to the creation of a sustainable business model [34]. Oderanti and Li presented a conceptual framework for a sustainable business involving assisted living technologies that included value proposition, product innovation and commercialization, infrastructure management, customer relation management, financial viability and sustainability, stakeholder credibility, and revenue streams as the 7 components. The comparative advantage of this framework was the most important factor that encouraged older people to pay for eHealth services, even though health services were free [35].

In 2 other studies, the new business models were slightly modified [30,36]. In Pereira's study, the 5-component VISOR interactive business model was the same as that presented in Fife and Pereira's study [30], but it was presented in more detail. This model had 5 components, namely value proposition, interface, service platform, organizing model, and revenue, and the results of the study showed that the weakness of one component could be compensated by strengthening another component [1].

Arkwright et al presented 8 common and successful telehealth business models in their study. These models included the direct-to-consumer (patient) business model, organization-to-organization business model, clinician-to-clinician business model, oversight and processes business model, online patient access/portals/technology, a business model based on mHealth/medical applications, a hardware/software model, and an international business model. The researchers believed that a successful telehealth business model should be safe, patient-centered, user-friendly, consistent, mission-oriented, strategy-oriented, and of proven value to the patient [36]. The key aspects of the aforementioned business models are presented in Figure 2.

Figure 2. Key aspects of the new business models or frameworks for the telehealth industry.



Existing Business models or Frameworks Used in the Telehealth Industry

The findings of this study showed that 10 studies examined existing business models or frameworks used in the telehealth industry (Table 2) [37-46]. Among them, 5 studies used the 9-component Osterwalder business model, which includes customer segments, value propositions, channels, customer relationships, key resources, key activities, key partnerships, cost structures, and revenue streams [37-40,43]. Hidefjäll and Titkova showed that the development of wearable sensor technologies should be considered as part of a more extensive commercialization process consisting of conceptual, financial, and organizational developments, and the requirements of the health system should be considered [37].

The results of Marjomaa's study [38] showed that the eHealth service market for chronic diseases was a multidisciplinary

market with several different segments, and the use of a participatory strategy such as the Osterwalder business model had a significant impact on the success of this market. Leeuwen has suggested that cost-benefit studies are essential for the success of assisted living technologies in dementia care, and they can be considered along with the components of the Osterwalder model [39]. Similarly, the results of a study conducted by Kho et al showed that although the Osterwalder model can be considered as a basis for different types of telehealth businesses, the 3 components of physician participation, medical risk management, and country-specific commitments must be considered to support the sustainability of teledermoscopy services [40]. Grustam et al have stated that although attention has been paid to the components of the Osterwalder model, synergy among manufacturers, health care providers, payers, and legislators is necessary to implement telescreening technology for patients with heart diseases [43].

Table 2. Summary of the studies that evaluated existing business models or frameworks used in the telehealth industry.

Author(s) (year) and country	Objective	Methods	Business model	Model components	Results
Dijkstra et al (2006) [44] Netherlands	To provide a business model for telemonitoring	Qualitative study (literature review)	Freeband business blueprint method	Four components: service domain, technological domain, organizational domain, and financial domain	Using 1 flexible infrastructure for multiple telemonitoring services, infrastructure costs can be shared among multiple services. A partnership between home care organizations, central contact centers, suppliers of monitoring devices, and wireless sensor network providers is required for telemonitoring.
Leunissen (2008) [45] Netherlands	To validate the process of the business model design of Myotel (see Table 1)	Qualitative study (case study)	STOF ^a model	Four components: service domain, technological domain, organizational domain, and financial domain	The business model is sustainable (viable, suitable for growth, and sustainable), if it has added value for all stakeholders involved.
Simonse et al (2011) [46] China	To review business models in home health services	Qualitative study (literature review)	Johnson framework	Four components: customer value proposition, profit formula, key resources, and key processes	There is an imbalance as to where money can be earned, where money can be saved, and where other value is created. Home health care providers are delivering extended, preventive, or outsourced health care from hospitals.
Marjomaa (2015) [38] Australia	To develop a generalizable business model for mHealth ^b services in chronic disease management	Mixed methods study (quantitative and qualitative)	Alexander Osterwalder's Business Model Canvas	Nine components: customer segments, value propositions, channels, customer relationships, revenue streams, key resources, key activities, key partnerships, and cost structure	Focusing on business model design early in the mHealth technology development phase can help researchers and designers overcome common challenges and create commercially viable mHealth services.
Hidefjäll and Titkova (2015) [37] Sweden	To design a business model for a wearable biofeedback system	Qualitative study (literature review and interviews with relevant representatives)	Alexander Osterwalder's Business Model Canvas	Nine components: customer segments, value propositions, channels, customer relationships, revenue streams, key resources, key activities, key partnerships, and cost structure	Instead of solely focusing on the material development of the technology, development needs to be seen as part of a larger commercialization process consisting of conceptual, material, and institutional development with the business model design in focus to meet health care system requirements.
Grustam et al (2017) [42] Netherlands	To assess the B2C ^c model for telemonitoring patients with chronic heart failure	Qualitative study (literature review)	B2C	Three components: design, structure, and governance	The B2C model in telemonitoring chronic heart failure potentially creates value for patients, who are shareholders of the service. Moreover, implementation of telemonitoring for chronic diseases via the B2C model can potentially free up financial resources, which can either be used to support a greater number of people with the same technology or can be invested in new treatments and therapies.
Grustam et al (2017) [41] Netherlands	To create the B2B ^d and B2C care models and explore the differences in care coordination and transaction costs between these models for telemonitoring	Qualitative study (literature review)	B2B and B2C	Six components: structure, financing, public policies, technology alignment, consumers (customers), and accountability	In principle, care coordination in the B2B and B2C models for telemonitoring chronic diseases differs in terms of design elements and design themes. The transaction costs could potentially be lower in the B2C model than in the B2B model, which could be a promoting economic principle.

Author(s) (year) and country	Objective	Methods	Business model	Model components	Results
Grustam et al (2018) [43] Netherlands	To describe a B2C model for the implementation of telemonitoring, by extending the current B2B model	Qualitative study (literature review)	Alexander Osterwalder's Business Model Canvas	Nine components: customer segments, value propositions, channels, customer relationships, revenue streams, key resources, key activities, key partnerships, and cost structures	A B2B model was developed toward a B2C model offered in telemonitoring with the goal of synergizing equipment manufacturers, health care providers, payers, and legislators to enable telemonitoring for the entire population and increase the speed and scalability of the technology.
Leeuwenwerden (2018) [39] Netherlands	To increase the commercial viability of business model innovations with SHAAL ^e technology in dementia care	Mixed methods study (quantitative and qualitative)	Alexander Osterwalder's Business Model Canvas	Nine components: customer segments, value propositions, channels, customer relationships, revenue streams, key resources, key activities, key partnerships, and cost structures	Cost-benefit studies were essential to the success of ambient assisted living technology, and the insurance company played an important role in continuing to use and commercialize these technologies.
Kho et al (2020) [40] Australia	To identify, describe, compare, and contrast the building blocks for direct-to-consumer mobile teledermoscopy services	Qualitative study (literature review)	Alexander Osterwalder's Business Model Canvas and Ash Maurya's Lean Canvas	Nine components of Alexander Osterwalder's business model canvas: customer segments, value propositions, channels, customer relationships, revenue streams, key resources, key activities, key partnerships, and cost structures Nine components of Ash Maurya's Lean Canvas: customer segments, problem, revenue streams, solution, unique value proposition, channels, key metrics, cost structure, and unfair advantage	The 3 business elements that supported the viability, sustainability, and growth of web-based dermatology were developing key partnerships, clinician involvement in the design and implementation process, and managing the medicolegal risks and liabilities that are relevant for each country.

^aSTOF: service, technological, organizational, and financial.

^bmHealth: mobile health.

^cB2C: Business-to-Consumer.

^dB2B: Business-to-Business.

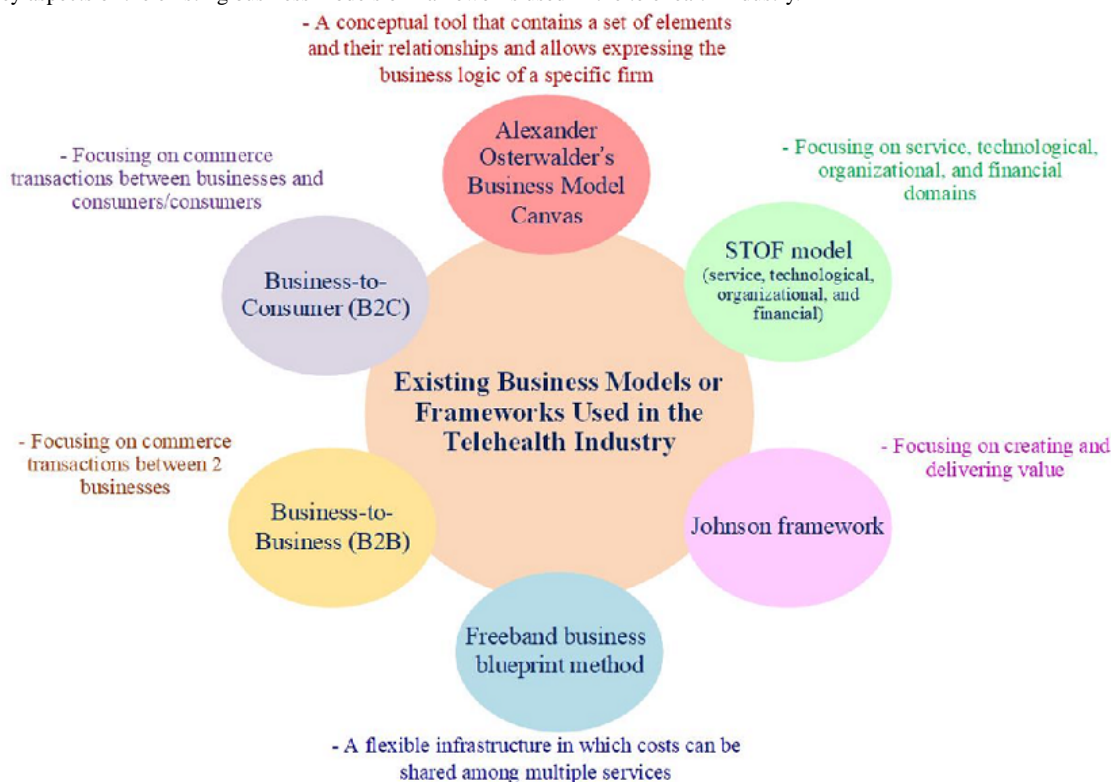
^eSHAAL: Smart Home and Ambient Assisted Living.

Business-to-Business (B2B) and Business-to-Consumer (B2C) models were used in 2 studies. The aim of 1 study was to explore the systemic and economic differences in care coordination via B2B and B2C models for telemonitoring patients with chronic diseases [41], and in another study, the aim was to assess the B2C model for telemonitoring patients with chronic heart failure by analyzing its value for organizations or ventures that provided telemonitoring services [42]. In these studies, the B2C model was used with its 6 components of structure, financing, public policies, technology alignment, consumers (customers), and accountability. This model created value for customers, shareholders, service providers, and the community [41,42].

Furthermore, 3 studies used other existing business models [44-46]. Dijkstra et al used the freeband business blueprint method (FBBM) including service domain, technological

domain, organizational domain, and financial domain as the components. The results indicated that costs can be divided between several telemonitoring services using a flexible infrastructure [44].

In a study conducted by Leunissen, the STOF (service, technological, organizational, and financial) model was used. The results showed that the added value in the telerehabilitation business model might be changed due to the impact of cash flows [45]. In another study, Simonse et al used the Johnson framework, which included customer value proposition, profit formula, key resources, and key processes. They noted that designing a business model is not separate from the organizational context [46]. The key aspects of the existing business models or frameworks used in the telehealth industry are illustrated in Figure 3.

Figure 3. Key aspects of the existing business models or frameworks used in the telehealth industry.

Synthesis

The results showed that different types of business models and frameworks have been used in the telehealth industry and they have various components. However, value proposition, meeting the stakeholders' and customers' requirements, and financial issues were the most common components in these models and frameworks. These components might be described using different terms along with many other components, which were found important in relation to a specific type of technology and its context of use. Although new business models and frameworks focus on specific aspects of telehealth services, namely, service delivery, innovation, technology, and interface design along with other business components, the existing business models, such as Osterwalder's business model canvas, have been used by some researchers to gain more comprehensive insight into the telehealth industry. It seems that using these business models and frameworks depends on the context of using the technology and many other components can be added to make them more appropriate for different purposes.

Discussion

Principal Findings

The aim of this study was to review different types of business models and their components used in the telehealth industry. The search process yielded 4998 articles, from which 23 studies were included in the study. These studies were divided into 2 main categories. The first category included new business models or frameworks, and the second category included the existing business models and frameworks used in telehealth industry. These models and frameworks consisted of different components in various areas of telehealth.

As mentioned earlier, business models can help implement telehealth technology with the participation of all stakeholders and in a value-based manner [4]. Business models serve as an analytical framework for identifying and overcoming barriers to the implementation and extensive use of telehealth technologies and help apply beneficial emerging technologies. These models also help identify the value proposition of telehealth services and its challenges, as well as the appropriate revenue model, organizational structure, and stakeholder engagement model [1]. However, business models must be adapted to the social, geographical, and economic contexts of the technology. Understanding each component of a business model is essential to evaluate the success of telehealth services [13,47]. Moreover, providing a business plan based on the well-known business models or frameworks, especially in the early stages of product development, will reduce potential risks and significantly save the costs related to the establishment of services and technologies [14,29,31].

A business model should be able to create and transfer value to the customers in a profitable and sustainable manner [23,30]. Therefore, some studies have emphasized the differences among the business models used for various types of telehealth technologies in each country [13,48]. For example, the results of the study conducted by Fredriksson et al showed that it is more appropriate to use different business frameworks for specific purposes. These frameworks should be in line with the context and purpose of using the technology [20]. However, the application of business models in the field of telehealth does not guarantee the success of new technologies, and before taking any action, legal issues and challenges related to licensing, compensation methods, liability, data sharing, and data protection must be resolved [28].

According to research findings, the main components in most telehealth business models were financial issues and cost structures that could be influenced by service processes, resources, and partners [33]. Cost structure plays an important role in customer acceptance, and different financial strategies need to be considered for various circumstances, revenue makings, and geographical areas [33]. Thus, a successful business model must be able to provide the highest value and increase the customers' willingness to pay [1,36].

The results showed that it is possible to design different types of business models with various components to be used in telehealth industry. However, the components should be able to support the value of the technology in line with other components, such as the cost structure and revenue model. The components of a business model must be able to support each other, especially in unstable conditions of the health system. In addition, the components of a business model must be constantly monitored and updated [34].

A business model should ultimately lead to the acceptance of the technology by the general public. It should help in providing equitable distribution of services, effective diagnosis of diseases, and high quality of services, as well as in reducing pressure on health systems [35].

The results also revealed that some studies used existing business models or frameworks in telehealth services. Among these models, the Osterwalder business model was used more frequently than other models [37-46]. This model was helpful to meet the requirements of the health system and provided added value by increasing patient satisfaction and reducing the cost of care [37,43]. It also provided a better understanding of the business characteristics and covered various economic aspects of technology implementation [40].

A number of other studies used B2C and B2B models. The use of the B2C model allowed all stakeholders to enjoy the benefits of innovation, reduced the burden of service delivery, and improved efficiency [42]. However, when insurance companies supported the B2B model, it was more sustainable than the B2C model. Other studies used the FBBM, Johnson framework, and STOF model. The use of these models was influenced by cash flow to generate revenue and predict outcomes [44-46]. Similarly, Antoniotti et al showed that government and private payers are very influential in making telehealth payments and revenue policies should be considered in business models [49].

Although business sustainability is one of the major challenges lying ahead for the expanding telehealth industry, few studies have concentrated on this aspect [32,34,35]. The key aspects of the long-term sustainability of telehealth business include developing a skilled workforce, empowering consumers, reforming funding, improving digital ecosystems, and integrating telehealth into routine care. These requirements should be considered in implementation planning to ensure that effective integration of telehealth within complex health systems is in place and staff are willing to use telehealth technologies [50,51]. In another study, Cui et al highlighted that the sustainability of telemedicine must be improved by appropriate legislation, uniform standards, and powerful management [52].

The existing business models, especially the Osterwalder business model, are general tools and roadmaps that can provide a good understanding of business model components. However, one of its major drawbacks is the lack of sufficient emphasis on the importance of the digital economy and the functionalities of core enabling technologies. In fact, this business model cannot manage multiservice platforms and the use of other business models seems necessary to support it. Moreover, it is more product-oriented and the nature of key partner networks is less discussed in this business model [38]. In situations where the stakeholders, their roles, and the impact of their roles are different, the existing business models do not have the necessary flexibility for adaptation. Moreover, the customers (organization purchasing technology) and users (technology user) are different sometimes and considering the requirements of both groups may influence the design of the business model [53].

Practical Recommendations

Overall, applying business models in the commercialization of telehealth services will be useful to gain a better understanding of the required components, stakeholders' interactions, market challenges, and possible future changes. In fact, understanding the innovation, market size, competitive strategy, and investment in the telehealth industry is not sufficient and the impact of such an investment on the whole society should be investigated [38,42]. Although several business models have been proposed for use in the telehealth industry, using a combination of models and their components can help commercialize the technology more successfully. Telehealth business models can also be used in combination with traditional patient care models to double their value proposition [40,43].

Strengths and Limitations

In this study, different types of business models and their components used in the telehealth industry were reviewed and the main components necessary for a successful telehealth business were identified. However, this study has some limitations. In most of the selected studies, qualitative approaches were used. Therefore, conducting meta-analysis was not possible. Moreover, although the main databases were searched, there might be other databases that were not searched and non-English papers that were excluded from the study. These limitations can be addressed in future studies by searching more databases and changing the exclusion criteria.

Conclusions

The results showed that new business models used in the telehealth industry focused on legal, organizational, insurance, and customer-related issues. Added value, financial variables, financial sustainability in the market, competition, service platform, annual membership and subscription, national incentives, cost structure, and revenue streams were the other important components of these models. The studies that used existing business models mostly focused on aspects such as design, structure, governance, organizational issues, country-specific obligations, public policy, financing, profit formula, physician participation, risk management, and key processes.

In general, the diversity of business models and their components in the telehealth industry indicates that different models can be used for different telehealth technologies in various health systems and cultures. However, it is necessary to evaluate the effectiveness of these models in practice. Moreover, comparing the usefulness of these models in different domains of telehealth services will help identify the strengths and weaknesses of these models for future optimization.

Acknowledgments

This work was funded and supported by Iran University of Medical Sciences, Tehran, Iran (grant IUMS/SHMIS_98-1-37-14528).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Search strategies.

[[DOCX File , 13 KB - jmir_v24i3e33128_app1.docx](#)]

Multimedia Appendix 2

CASP (Critical Appraisal Skills Programme) checklist.

[[DOCX File , 37 KB - jmir_v24i3e33128_app2.docx](#)]

References

- Pereira F. Business models for telehealth in the US: analyses and insights. *Smart Homecare Technol Telehealth* 2017 Feb;4:13-29. [doi: [10.2147/shtt.s68090](#)]
- Yang YT, Lin NP, Su S, Chen YM, Chang YM, Handa Y, et al. Value-added strategy models to provide quality services in senior health business. *Int J Qual Health Care* 2017 Jun;32(1):1-5. [doi: [10.1093/intqhc/mzx072](#)] [Medline: [28637191](#)]
- Saigí F, Sellens J, Ramos I, Sáez C, Kotzeva A, Hidalgo J. Framework for the implementation of a telemedicine service. *Rev Eur Comp Int Environ Law* 2016;25(2):228-237.
- van Limurg M, van Gemert-Pimjen JEW, Nijland N, Ossebaard HC, Hendrix RMG, Seydel ER. Why business modeling is crucial in the development of eHealth technologies. *J Med Internet Res* 2011 Oct;13(4):e124 [FREE Full text] [doi: [10.2196/jmir.1674](#)] [Medline: [22204896](#)]
- Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: opportunities and barriers. *Am J Med* 2014 Mar;127(3):183-187. [doi: [10.1016/j.amjmed.2013.09.032](#)] [Medline: [24384059](#)]
- Seelman KD, Hartman LM. Telerehabilitation: policy issues and research tools. *Int J Telerehabil* 2009 Sep;1(1):47-58 [FREE Full text] [doi: [10.5195/ijt.2009.6013](#)] [Medline: [25945162](#)]
- Cottrell MA, Russell TG. Telehealth for musculoskeletal physiotherapy. *Musculoskelet Sci Pract* 2020 Aug;48:102193 [FREE Full text] [doi: [10.1016/j.msksp.2020.102193](#)] [Medline: [32560876](#)]
- van Dyk L. A review of telehealth service implementation frameworks. *Int J Environ Res Public Health* 2014 Jan;11(2):1279-1298 [FREE Full text] [doi: [10.3390/ijerph110201279](#)] [Medline: [24464237](#)]
- Weidner K, Lowman J. Telepractice for adult speech-language pathology services: a systematic review. *Perspect ASHA SIGs* 2020 Feb;5(1):326-338. [doi: [10.1044/2019_PERSP-19-00146](#)]
- Orlando JF, Beard M, Kumar S. Systematic review of patient and caregivers' satisfaction with telehealth videoconferencing as a mode of service delivery in managing patients' health. *PLoS One* 2019 Aug;14(8):e0221848 [FREE Full text] [doi: [10.1371/journal.pone.0221848](#)] [Medline: [31469865](#)]
- Becevic M, Sheets LR, Wallach E, McEowen A, Bass A, Mutrux ER, et al. Telehealth and Telemedicine in Missouri. *Mo Med* 2020 May;117(3):228-234 [FREE Full text] [Medline: [32636555](#)]
- Harst L, Lantzsch H, Scheibe M. Theories predicting end-user acceptance of telemedicine use: systematic review. *J Med Internet Res* 2019 May;21(5):e13117 [FREE Full text] [doi: [10.2196/13117](#)] [Medline: [31115340](#)]
- Chen S, Cheng A, Mehta K. A review of telemedicine business models. *Telemed J E Health* 2013 Mar;19(4):287-297. [doi: [10.1089/tmj.2012.0172](#)] [Medline: [23540278](#)]
- Kijl B, Nieuwenhuis LJM, Huis in 't Veld RMHA, Hermens HJ, Vollenbroek-Hutten MMR. Deployment of e-health services—a business model engineering strategy. *J Telemed Telecare* 2010 Aug;16(6):344-353. [doi: [10.1258/jtt.2010.006009](#)] [Medline: [20798429](#)]
- Kijl B, Nieuwenhuis B. Deploying a telerehabilitation service innovation: an early stage business model engineering approach. In: *Annual Hawaii International Conference on System Sciences (HICSS)*. 2010 Presented at: 2010 43rd Hawaii International Conference on System Sciences; January 5-8, 2010; United States p. 1-10. [doi: [10.1109/HICSS.2010.134](#)]

16. Kimble C. Business models for e-health: evidence from ten case studies. *Glob Bus Organ Excell* 2015 May;34(4):18-30. [doi: [10.1002/joe.21611](https://doi.org/10.1002/joe.21611)]
17. Lehoux P, Daudelin G, Williams-Jones B, Denis J, Longo C. How do business model and health technology design influence each other? Insights from a longitudinal case study of three academic spin-offs. *Research Policy* 2014 Jul;43(6):1025-1038. [doi: [10.1016/j.respol.2014.02.001](https://doi.org/10.1016/j.respol.2014.02.001)]
18. Kluge EHW. Ethical and legal challenges for health telematics in a global world: telehealth and the technological imperative. *Int J Med Inform* 2011 Feb;80(2):e1-e5. [doi: [10.1016/j.ijmedinf.2010.10.002](https://doi.org/10.1016/j.ijmedinf.2010.10.002)]
19. Peters C, Blohm I, Leimeister JM. Anatomy of successful business models for complex services: insights from the telemedicine field. *J Manag Inform Syst* 2015 Dec;32(3):75-104. [doi: [10.1080/07421222.2015.1095034](https://doi.org/10.1080/07421222.2015.1095034)]
20. Fredriksson JJ, Mazzocato P, Muhammed R, Savage C. Business model framework applications in health care: a systematic review. *Health Serv Manage Res* 2017 Nov;30(4):219-226. [doi: [10.1177/0951484817726918](https://doi.org/10.1177/0951484817726918)] [Medline: [28868934](https://pubmed.ncbi.nlm.nih.gov/28868934/)]
21. Acheampong F, Vimarlund V. Business models for telemedicine services: a literature review. *Health Syst* 2015 Nov;4(3):189-203. [doi: [10.1057/hs.2014.20](https://doi.org/10.1057/hs.2014.20)]
22. Meertens LO, Iacob ME, Nieuwenhuis JM. A method for business model development. In: *Proceedings of International Symposium on Business Modeling and Software Design*. Germany: Springer; 2012 Presented at: International Symposium on Business Modeling and Software Design; July 27-28, 2011; Bulgaria p. 113-129.
23. Barker GP, Krupinski EA, McNeely RA, Holcomb MJ, Lopez AM, Weinstein RS. The Arizona Telemedicine Program business model. *J Telemed Telecare* 2005 Dec;11(8):397-402. [doi: [10.1177/1357633X0501100804](https://doi.org/10.1177/1357633X0501100804)] [Medline: [16356313](https://pubmed.ncbi.nlm.nih.gov/16356313/)]
24. Tawfik GM, Dila KAS, Mohamed MYF, Tam DNH, Kien ND, Ahmed AM, et al. A step by step guide for conducting a systematic review and meta-analysis with simulation data. *Trop Med Int Health* 2019 Aug;47:46 [FREE Full text] [doi: [10.1186/s41182-019-0165-6](https://doi.org/10.1186/s41182-019-0165-6)] [Medline: [31388330](https://pubmed.ncbi.nlm.nih.gov/31388330/)]
25. 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;4:1-9 [FREE Full text] [doi: [10.1186/2046-4053-4-1](https://doi.org/10.1186/2046-4053-4-1)] [Medline: [25554246](https://pubmed.ncbi.nlm.nih.gov/25554246/)]
26. 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;6(7):e1000097 [FREE Full text] [doi: [10.1371/journal.pmed.1000097](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
27. Ma L, Wang Y, Yang Z, Huang D, Weng H, Zeng X. Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: what are they and which is better? *Mil Med Res* 2020 Feb;7:7. [doi: [10.1186/s40779-020-00238-8](https://doi.org/10.1186/s40779-020-00238-8)] [Medline: [32111253](https://pubmed.ncbi.nlm.nih.gov/32111253/)]
28. Mun SK, Tohme WG, Platenberg RC, Choi I. Teleradiology and emerging business models. *J Telemed Telecare* 2005;11(6):271-275. [doi: [10.1258/1357633054893373](https://doi.org/10.1258/1357633054893373)] [Medline: [16168162](https://pubmed.ncbi.nlm.nih.gov/16168162/)]
29. Lee Y, Chang P. Modeling a mobile health management business model for chronic kidney disease. *Stud Health Technol Inform* 2016;225:1047-1048. [doi: [10.3233/978-1-61499-658-3-1047](https://doi.org/10.3233/978-1-61499-658-3-1047)] [Medline: [27332476](https://pubmed.ncbi.nlm.nih.gov/27332476/)]
30. Fife E, Pereira F, Sawcer D. The business models for mobile telehealth in the US: applying the VISOR framework. In: *Proceedings of the 7th Mobility Roundtable*. 2008 Presented at: The 7th Mobility Roundtable; 2008; New Zealand p. 1-11 URL: <https://docplayer.net/6737187-The-business-models-for-mobile-tele-health-in-the-u-s-applying-the-visor-framework.html>
31. Fusco F, Turchetti G. Interactive business models for telerehabilitation after total knee replacement. In: *Proceedings of International Conference on Bioinformatics and Biomedical Engineering*.: Springer; 2015 Presented at: International Conference on Bioinformatics and Biomedical Engineering; April 15-17, 2015; Spain p. 502-511 URL: <http://kt.ijs.si/interno/Proceedings/IWBBIO2015/papers/9044/90440049.pdf>
32. Lin S, Liu J, Wei J, Yin W, Chen H, Chiu W. A business model analysis of telecardiology service. *Telemed J E Health* 2010 Dec;16(10):1067-1073. [doi: [10.1089/tmj.2010.0059](https://doi.org/10.1089/tmj.2010.0059)] [Medline: [21186990](https://pubmed.ncbi.nlm.nih.gov/21186990/)]
33. Lin T, Chang H, Huang C. An analysis of telemedicine in Taiwan: a business model perspective. *Int J Gerontol* 2011 Dec;5(4):189-192. [doi: [10.1016/j.ijge.2011.09.039](https://doi.org/10.1016/j.ijge.2011.09.039)]
34. Fachinger U, Schöpke B. Business model for sensor-based fall recognition systems. *Inform Health Soc Ca* 2014 Aug;39(3-4):305-318. [doi: [10.3109/17538157.2014.931855](https://doi.org/10.3109/17538157.2014.931855)] [Medline: [25148563](https://pubmed.ncbi.nlm.nih.gov/25148563/)]
35. Oderanti FO, Li F. A holistic review and framework for sustainable business models for assisted living technologies and services. *Int J Healthc Tech Manag* 2017 May;15(4):273-307. [doi: [10.1504/ijhtm.2016.084128](https://doi.org/10.1504/ijhtm.2016.084128)]
36. Arkwright BT, Leslie M, Light M. Telehealth finance variables and successful business models. In: Ford DW, Valenta SR, editors. *Telemedicine. Respiratory Medicine*. Switzerland: Humana; 2021:33-59.
37. Hidefjäll P, Titkova D. Business model design for a wearable biofeedback system. *Stud Health Technol Inform* 2015;211:213-224. [Medline: [25980872](https://pubmed.ncbi.nlm.nih.gov/25980872/)]
38. Marjomaa S. Business model for mobile healthcare delivery in chronic disease management [dissertation]. Tampere University of Technology. 2015. URL: <https://trepo.tuni.fi/handle/123456789/23604> [accessed 2022-03-08]
39. Leeuwen M. Assessing the impact of smart home and ambient assisted living technology on the dementia care value network and business model [dissertation]. University of Twente Student Theses.: University of Twente; 2018. URL: <https://essay.utwente.nl/77002/> [accessed 2022-03-08]

40. Kho J, Gillespie N, Horsham C, Snoswell C, Vagenas D, Soyer HP, et al. Skin doctor consultations using mobile teledermoscopy: exploring virtual care business models. *Telemed J E Health* 2020 Nov;26(11):1406-1413. [doi: [10.1089/tmj.2019.0228](https://doi.org/10.1089/tmj.2019.0228)]
41. Grustam AS, Vrijhoef H, Cordella A, Koymans R, Severens JL. Care coordination in a business-to-business and a business-to-consumer model for telemonitoring patients with chronic diseases. *Int J Care Coord* 2017 Dec;20(4):135-147 [FREE Full text] [doi: [10.1177/2053434517747908](https://doi.org/10.1177/2053434517747908)] [Medline: [29276610](https://pubmed.ncbi.nlm.nih.gov/29276610/)]
42. Grustam AS, Vrijhoef HJM, Koymans R, Hukal P, Severens JL. Assessment of a Business-to-Consumer (B2C) model for telemonitoring patients with chronic heart failure (CHF). *BMC Med Inform Decis Mak* 2017 Oct;17:145 [FREE Full text] [doi: [10.1186/s12911-017-0541-2](https://doi.org/10.1186/s12911-017-0541-2)] [Medline: [29020993](https://pubmed.ncbi.nlm.nih.gov/29020993/)]
43. Grustam AS, Vrijhoef HJM, Poulidikis V, Koymans R, Severens JL. Extending the Business-to-Business (B2B) model towards a Business-to-Consumer (B2C) model for telemonitoring patients with chronic heart failure. *J Bus Models* 2018;6(3):106-129.
44. Dijkstra S, Jurriëns J, Mei R. A business model for telemonitoring services. In: *Proceedings of 14th Annual High Technology Small Firms Conference*. 2006 Presented at: 14th Annual High Technology Small Firms Conference; January 2006; Netherlands p. 1-12 URL: <http://proceedings.utwente.nl/145/1/Dijkstra%2CS.J.c.s.paper.pdf>
45. Leunissen R. The validation of an e-health service business model. In: *Proceedings of 10th Twente Student Conference on IT*. 2008 Presented at: 10th Twente Student Conference on IT; 2008; Netherlands.
46. Simonsen W, Zonneland J, Liu K, Govers F, Vincent R, Laban V. Design research on business models in home healthcare. In: *Proceedings of Tsinghua-DMI: Tsinghua-DMI*; 2011 Presented at: International Design Management Symposium; December 3-5, 2011; Hong Kong URL: <http://resolver.tudelft.nl/uuid:e827dc6e-1020-4205-9b58-754a36454f66>
47. Dargahi H, Mohammadzadeh N, Rezaeyanzadeh H. E-business in health care systems. *Payavard Salamat* 2012;5(3):39-48.
48. Gorski I, Bram JT, Sutermeister S, Eckman M, Mehta K. Value propositions of mHealth projects. *J Med Eng Technol* 2016 Aug;40(7-8):400-421. [doi: [10.1080/03091902.2016.1213907](https://doi.org/10.1080/03091902.2016.1213907)]
49. Antoniotti NM, Drude KP, Rowe N. Private payer telehealth reimbursement in the United States. *Telemed J E Health* 2014 Jun;20(6):539-543. [doi: [10.1089/tmj.2013.0256](https://doi.org/10.1089/tmj.2013.0256)] [Medline: [24654748](https://pubmed.ncbi.nlm.nih.gov/24654748/)]
50. Thomas EE, Haydon HM, Mehrotra A, Caffery LJ, Snoswell CL, Banbury A, et al. Building on the momentum: sustaining telehealth beyond COVID-19. *J Telemed Telecare* 2020 Sep;1357633X20960638. [doi: [10.1177/1357633X20960638](https://doi.org/10.1177/1357633X20960638)] [Medline: [32985380](https://pubmed.ncbi.nlm.nih.gov/32985380/)]
51. Velayati F, Ayatollahi H, Hemmat M, Dehghan R. Key components and critical factors for developing a telehealth business framework: a qualitative study. *BMC Med Inform Decis Mak* 2021 Dec;21:339. [doi: [10.1186/s12911-021-01707-3](https://doi.org/10.1186/s12911-021-01707-3)]
52. Cui F, Ma Q, He X, Zhai Y, Zhao J, Chen B, et al. Implementation and application of telemedicine in China: cross-sectional study. *JMIR Mhealth Uhealth* 2020 Oct;8(10):e18426 [FREE Full text] [doi: [10.2196/18426](https://doi.org/10.2196/18426)] [Medline: [33095175](https://pubmed.ncbi.nlm.nih.gov/33095175/)]
53. Nikou S, Bouwman H. Mobile health and wellness applications: a business model ontology-based review. *Int J E-Bus Res* 2017;13(1):1-24. [doi: [10.4018/IJEBR.2017010101](https://doi.org/10.4018/IJEBR.2017010101)]

Abbreviations

B2B: Business-to-Business

B2C: Business-to-Consumer

CASP: Critical Appraisal Skills Programme

mhealth: mobile health

FBBM: freeband business blueprint method

STOF: service, technology, organization, finance

VISOR: value proposition, interface, service platform, organizing model, and revenue

Edited by R Kukafka; submitted 24.08.21; peer-reviewed by M Arab, V Strotbaum; comments to author 16.09.21; revised version received 31.10.21; accepted 15.11.21; published 29.03.22.

Please cite as:

Velayati F, Ayatollahi H, Hemmat M, Dehghan R

Telehealth Business Models and Their Components: Systematic Review

J Med Internet Res 2022;24(3):e33128

URL: <https://www.jmir.org/2022/3/e33128>

doi: [10.2196/33128](https://doi.org/10.2196/33128)

PMID: [35348471](https://pubmed.ncbi.nlm.nih.gov/35348471/)

©Farnia Velayati, Haleh Ayatollahi, Morteza Hemmat, Reza Dehghan. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 29.03.2022. 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.

Review

Synthesizing Dimensions of Digital Maturity in Hospitals: Systematic Review

Rhona Duncan¹, BIT; Rebekah Eden¹, BIT, BAPSC, PhD; Leanna Woods^{2,3,4}, BNurs, PGCert, PhD; Ides Wong⁵, BBehSc, PhD; Clair Sullivan^{2,4,6}, MBBS, MD

¹School of Information Systems, Queensland University of Technology, Brisbane, Australia

²Centre for Health Services Research, The University of Queensland, Herston, Australia

³Digital Health Cooperative Research Centre, Australian Government, Sydney, Australia

⁴Digital Health Research Network, The University of Queensland, Brisbane, Australia

⁵Clinical Excellence Queensland, Queensland Health, Brisbane, Australia

⁶Metro North Hospital and Health Service, Brisbane, Australia

Corresponding Author:

Rebekah Eden, BIT, BAPSC, PhD

School of Information Systems

Queensland University of Technology

2 George Street

Brisbane, 4000

Australia

Phone: 61 434237975

Email: rg.eden@qut.edu.au

Abstract

Background: Digital health in hospital settings is viewed as a panacea for achieving the “quadruple aim” of health care, yet the outcomes have been largely inconclusive. To optimize digital health outcomes, a strategic approach is necessary, requiring digital maturity assessments. However, current approaches to assessing digital maturity have been largely insufficient, with uncertainty surrounding the dimensions to assess.

Objective: The aim of this study was to identify the current dimensions used to assess the digital maturity of hospitals.

Methods: A systematic literature review was conducted of peer-reviewed literature (published before December 2020) investigating maturity models used to assess the digital maturity of hospitals. A total of 29 relevant articles were retrieved, representing 27 distinct maturity models. The articles were inductively analyzed, and the maturity model dimensions were extracted and consolidated into a maturity model framework.

Results: The consolidated maturity model framework consisted of 7 dimensions: strategy; information technology capability; interoperability; governance and management; patient-centered care; people, skills, and behavior; and data analytics. These 7 dimensions can be evaluated based on 24 respective indicators.

Conclusions: The maturity model framework developed for this study can be used to assess digital maturity and identify areas for improvement.

(*J Med Internet Res* 2022;24(3):e32994) doi:[10.2196/32994](https://doi.org/10.2196/32994)

KEYWORDS

digital maturity; digital capability; eHealth; digital hospitals; capability model; maturity model; literature review; electronic medical records

Introduction

Planning a strategic roadmap to successful digital health transformation is challenging [1] due to a busy health landscape with competing drivers for change [2-4]. This is further compounded by the myriad of new technologies health care

providers can select from to advance their digital health agenda. Despite both the rapid global uptake of eHealth technologies [5] and digital health being viewed as a panacea [6] for achieving the “quadruple aim” of health care (ie, reducing costs, improving patient experience, improving the work life of health care providers, and advancing population health) [7], the

outcomes of digital health transformation are inconclusive and mixed [8,9]. One proposed method for strategically developing a digital health agenda is to follow a roadmap informed by digital maturity assessments [10,11].

In health care, digital maturity is defined as the extent to which digital systems are leveraged to provide high-quality health care, resulting in improved services and service delivery for an enhanced patient experience [12]. Assessing digital maturity is particularly important in hospital settings, due to (1) the complexity and cost of health service delivery involving multidisciplinary teams in acute, high-cost care settings [13]; (2) the necessity for rapid digital transformation that leverages eHealth technologies to cater to the needs of an aging population with increased rates of chronic disease [14]; and (3) the difficulties justifying business cases for large-scale electronic medical record system implementations, which require significant upfront and ongoing costs [15].

To assess digital maturity, a maturity model can be used to allow an organization to evaluate its current digital status across a series of dimensions [1]. However, limitations exist in current approaches for measuring digital maturity in hospitals, as there is a lack of consensus over which dimensions should be assessed [11]. Others have argued that current assessments of digital maturity are insufficient due to their primary focus on technology, with limited incorporation of organizational and human factors [4]. This is further supported by Carvalho et al [16], who emphasize that most digital maturity models lack sufficient depth and breadth for adequate assessment. Currently, there is still no agreement or convergence on how to assess digital maturity in health care.

Failure to understand the appropriate dimensions for assessing the digital maturity of hospitals will hamper the success of digital health transformation and be detrimental to health care outcomes. Therefore, a systematic literature review was conducted to find what dimensions are currently used to assess digital maturity in hospitals. As such, our aim was to synthesize the maturity model dimensions that are currently used when assessing digital hospital maturity to develop a consolidated digital maturity framework. Such a synthesis is necessary for, and will be beneficial to, health care executives and strategic decision-makers in evaluating and planning for the transformation of their practice. In addition, this synthesis will be beneficial to researchers as it consolidates the maturity dimensions and provides areas for future research to further refine and strengthen maturity models and their applications.

Methods

A systematic literature review following the guidelines of Templier and Pare [17] was conducted in December 2020 of articles that describe how digital maturity is assessed in hospital settings. In line with these guidelines, this systematic literature review was developmental in nature, in that it sought to develop a consolidated digital maturity framework. Therefore, unlike aggregative reviews, which seek to include *all* articles relevant to the phenomenon of interest, developmental reviews seek to cover only a *sample* of articles relevant to the phenomenon of interest [17].

To extract articles, medical databases (eg, PubMed, Cochrane, and Medline) and the Association for Information Systems College of Senior Scholars' Basket of Eight journals were searched using the following search string: (“maturity model*” OR “digital capabilit*” OR “digital maturity”). These databases and sources were selected due to the prominence of digital health in these domains. However, due to the breadth of information systems literature examining digital maturity across a myriad of contexts other than health care, the following search condition was added for a more targeted review of the information systems sources: “AND (“health” OR “healthcare”)”. This additional search condition was not added to medical databases due to their targeted focus.

Articles were excluded if they were (1) focused on settings other than hospitals, as the implementation of eHealth technologies in different contexts (eg, acute vs primary care) requires vastly different resources with large heterogeneity in impact measurement; (2) focused on maturity models not related to digital health (ie, training maturity); (3) not focused on digital maturity; (4) published in a language other than English; and (5) not a full-text article (ie, posters or extended abstracts).

As illustrated in Figure 1 [18], 357 articles were returned from the search, and after removing duplicates, 215 remained. Initially, the first author screened the abstracts resulting in 149 articles being removed and 66 potentially relevant articles being retained. Next, a full-text review was conducted to determine eligibility, resulting in the exclusion of 37 articles. A total of 29 articles were deemed applicable for analysis. To ensure no relevant articles were missed, backward searching of the references was performed. Consistent with Saldaña [19], intercoder corroboration was performed at each stage by the second author when determining whether an article should be included.

To analyze the relevant articles, inductive coding [20] was performed in NVivo (version 12 Plus; QSR International), with maturity model dimensions extracted. These were first extracted using verbatim codes [19] with 245 raw maturity nodes (the nodes included terms such as digital architecture, enterprise architecture, infrastructure, technology capabilities, reliability, decision support systems, picture archiving and communication system [PACS], and software applications). In some instances, the raw maturity nodes represented the specific maturity model dimensions as mentioned in the papers, while in other instances it referred to digital maturity stages, as some maturity models only provided stages rather than specific dimensions.

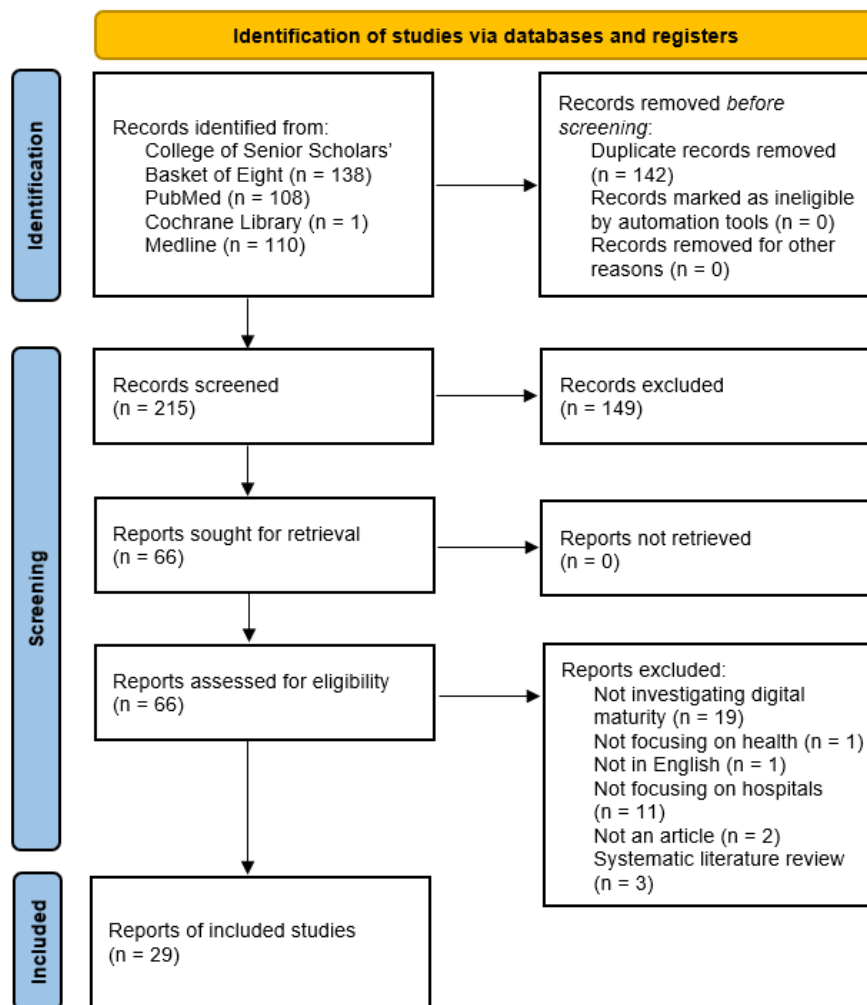
Through the constant comparison method [20], these raw maturity nodes were grouped into respective indicators based on commonalities. This involved considering the definition of each raw maturity node, as in some cases the name of the raw maturity node (as extracted verbatim from the paper) did not reflect its inherent meaning. By comparing the specific definitions of the raw maturity nodes, similar definitions were consolidated into a single indicator. For instance, digital architecture, enterprise architecture, and infrastructure were all related to information technology (IT), so were grouped into the IT infrastructure indicator, while technology capabilities and system reliability were grouped into the technical quality

indicator, and decision support systems, PACS, and software applications were grouped into the systems and services indicator. In total, 24 indicators were identified. Following this, constant comparison was again performed to aggregate the indicators into a consolidated set of dimensions based on commonalities amongst indicators. For instance, the IT infrastructure, technical quality, and systems and services indicators were grouped into the IT capability dimension. In total, 7 dimensions were identified.

To provide further confidence in our findings, reliability assessments were also performed. First, coder corroboration was conducted. The first author independently performed coding of verbatim measures (ie, raw maturity nodes) to indicators, then grouped the indicators into dimensions and discussed these decisions with the second author until consensus was reached

[19]. This involved ensuring that all verbatim measures were accurately mapped to the indicators. Through discussion, some of the verbatim measures were moved to a different indicator to better reflect their underlying definitions. Subsequently, additional coder corroboration of the indicators and dimensions was performed. This involved the first 2 authors discussing the indicators and dimensions with the rest of the authorship team and resulted in updates to some of the names and definitions [19]. Second, external reliability checks were performed and the dimensions were discussed at external forums, including (1) a statewide digital health steering committee in April 2021, attended by 14 members, and (2) the statewide “Digital Health Grand Rounds” in May 2021, attended by 120 health service executives, digital health researchers, and clinicians. Consensus was reached at the forums on the derived dimensions.

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



Results

Digital Maturity Dimensions

In total, 27 distinct maturity models were examined (Multimedia Appendix 1). In some instances, multiple maturity models were examined in a single paper, and in other instances, the same maturity model was examined in multiple papers. In total, 14 papers validated an existing maturity model, 10 papers proposed a new maturity model but did not validate the maturity model,

4 papers both proposed and validated a new maturity model, and 1 paper extended an existing maturity model.

Overall, 24 indicators were identified, which were consolidated into 7 digital maturity dimensions: strategy; IT capability; interoperability; governance and management; patient-centered care; people, skills, and behavior; and data analytics. The dimensions are described in Table 1; detailed examples of the indicators are provided with examples of their verbatim measures in Multimedia Appendix 2. Multimedia Appendix 3

illustrates how the distinct digital maturity models mapped onto the 7 digital maturity dimensions.

The findings of this review identified that digital maturity is predominantly assessed based on management-oriented dimensions and technology-related dimensions. Governance

and management (n=22 articles) has been the most prevalent dimension of digital maturity, followed by IT capability (n=18), people, skills, and behavior (n=17), interoperability (n=15), and strategy (n=14). Comparatively, limited research has examined data analytics (n=6) and patient-centered care (n=3). Each dimension is further described in the below subsections.

Table 1. Description of the Digital Maturity Dimensions.

Dimension	Description	Indicators
Strategy	The extent to which the organization has developed and implemented a strategic plan to achieve its goals and objectives [16]	Strategic adaptability, strategic alignment, strategic focus
Information technology capability	The extent to which the organization has adopted and implemented information technology infrastructure, digital systems, technologies, and services [21] that are usable and effective [22]	Information technology infrastructure, technical quality, systems and services
Interoperability	The extent to which data and information can be exchanged between systems within the organization, across care settings, and with patients, caregivers, and families [11]	External interoperability, internal interoperability, semantic interoperability, syntactic interoperability
Governance and management	The extent to which the organization embraces leadership, policies and procedures, structures, risk management of quality and safety, integrated workflows, relationship building, and capacity building [23]	Change management, data governance, leadership and management, risk management, standards, cultural values
Patient-centered care	The extent to which patients, caregivers, and families can actively participate in their health decisions, have access to information and health data, and cocreate services and service delivery [24]	Patient empowerment, patient focus
People, skills, and behavior	The extent to which stakeholders (internal and external) are digitally literate and motivated to leverage technology [11,25]	Education and training, knowledge management, individual competence, technology usage
Data analytics	The extent to which the organization uses data for effective decision-making for the organization, patients, and population health [1]	Descriptive analytics, predictive analytics

Governance and Management

The governance and management dimension is described as the extent to which the health care organization possesses formalized and committed leadership, as well as formalized policies, procedures, structures, and workflows [23]. Six indicators comprise the governance and management dimension: leadership and management, change management, cultural values, standards, risk management, and data governance.

The leadership and management indicator encompasses the executive team's commitment to and support for improving clinical quality [26-28] and fostering innovation across the hospital [24]. This support is essential for all levels of the workforce. The change management indicator recognizes the need to encourage individuals to embrace planned change to achieve desired outcomes [11,29]. The need for innovation and embracing change is further evident in the cultural values indicator, which espouses values of encouraging innovative behaviors [11,29] within a trusting [26-28] and inclusive environment [30,31]. The standards indicator assesses the extent to which processes [30-32], policies, and procedures are based on the standards that have been formally agreed and mandated [24] and the extent to which these contribute to optimizing the health care organization [11,12,33]; nevertheless, this indicator is not contrary to innovation. The risk management indicator acknowledges the need for the workforce to identify, mitigate, and report risks to ensure the safety, security, and privacy of patients [1,11,26-28] and the workforce [21]. The data

governance indicator further assesses whether data integrity, security, and privacy are preserved across the digital systems in health care settings [12], supported by standardized processes and protocols for accessibility and authorization [22,23,34].

IT Capability

The IT capability dimension represents the extent to which the organization has implemented IT infrastructure, digital systems, technologies, and services [21] that are usable and effective [22]. This dimension comprised three indicators: systems and services, IT infrastructure, and technical quality.

The systems and services indicator, which examines the digital systems implemented to support clinical care, is the most prominent indicator within this dimension. The systems identified as being important to digital maturity include electronic medical records, clinical decision support systems, e-prescribing, PACS [11,35,36], orders and results management, asset and resource optimization systems [22], and remote and assistive care systems [1,21]. The IT infrastructure indicator focuses on infrastructure [21,37] and architecture [1] designed and installed to support the aforementioned systems and services [12]. The systems and services indicator, as well as the IT infrastructure indicator, largely examine what technology and supporting structures have been implemented but do not account for their effectiveness. Alternatively, the technical quality indicator focuses on how effective, efficient, and fit for purpose the digital systems are [21,38].

People, Skills, and Behavior

The people, skills, and behavior dimension assesses the extent to which stakeholders, both internal and external to the health care organization, are digitally literate and motivated to leverage digital health systems [11,25]. This dimension consists of four indicators: education and training, knowledge management, individual competence, and technology usage.

The education and training indicator relates to the strategies adopted by the health care organization to provide individuals with opportunities to grow and develop clinical and technical skills, as well as collaboration and teamwork skills [16,29]. The knowledge management indicator refers to the extent to which workforce capability grows through creating, managing, and sharing knowledge [23,30]. These two indicators focus on ensuring organizational practices are in place to foster skill development and knowledge acquisition, whereas the individual competence and technology usage indicators focus on the actual skill sets and behaviors of individuals. For instance, at the individual level, the individual competence indicator takes into consideration that individuals need to possess skills, knowledge, and capability to use digital systems [21]. In contrast, the technology usage indicator recognizes that systems can be used in different ways and that digitally mature organizations need to ensure systems are used as intended [12] in a pervasive and consistent manner [26-28].

Interoperability

The interoperability dimension represents the extent to which data and information can be exchanged between systems within the organization, across care settings, and with patients, caregivers, and families [11]. Four interoperability indicators were identified: external interoperability, internal interoperability, semantic interoperability, and syntactic interoperability. The former 2 indicators relate to how information is exchanged between different actors within and between organizations (ie, intraorganizational vs interorganizational information exchange). The latter 2 indicators relate to data transformation and distinguish between the technical and meaningful structure of the information exchanged.

The external interoperability indicator assesses the adoption of standards to integrate systems, services, and data across the entire health care system [1,11,30,39-41]. Conversely, the internal interoperability indicator assesses the integration of systems and data across departments within a single health care organization [30,40]. The external interoperability indicator was more prevalent in the literature than the internal interoperability indicator.

The semantic interoperability indicator examines the extent to which information exchanged between digital systems can be accurately interpreted and understood by each system involved [34,42]. As such, the semantic interoperability indicator is dependent on the transparency of the underlying lexicon and data dictionary to ensure the intended meaning of the information exchange is retained [42]. Alternatively, the syntactic interoperability indicator represents the extent to which technical standards have been defined to enable the consistent,

effective, and efficient integration of digital systems and services [30,34].

Strategy

The strategy dimension represents the extent to which the organization has developed and implemented a strategic plan to achieve its goals and objectives [16]. The strategy dimension includes three indicators: strategic focus, strategic alignment, and strategic adaptability. This dimension is built on the premise that the digital strategy and the organizational strategy should be aligned and adaptable to support the accomplishment of measurable goals and outcomes related to quality and safety [26-28].

The strategic focus indicator was the most prevalent, with an emphasis on quality and safety [26-28], sustainability and cost effectiveness [39], and ensuring the systematic evaluation of quantifiable results and objectives [24,26-28]. While the strategic focus indicator centers on the core elements that health care organizations focus on, the strategic alignment indicator details the need for the digital strategy to be aligned with the organizational strategy [12,43]. To accomplish this, the digital strategy needs to be grounded on clinical benefits and outcomes [1,11]. In contrast, the strategic adaptability indicator recognizes the importance of organizational strategy and digital system dynamism [1,11,16] and that both should be capable of responding to environmental challenges [44] and emerging opportunities [11,16,24].

Data Analytics

The data analytics dimension examines the extent to which the organization uses data collected in its digital systems for effective decision-making to benefit the organization, patients, and population health [1]. Few studies have reported on this dimension and as such only 2 indicators have been identified: descriptive analytics and predictive analytics.

The descriptive analytics indicator is the extent to which data is analyzed to identify and understand historical patterns and trends, facilitating effective decision-making [1]. The predictive analytics indicator focuses on the analysis of data that enables future potential risks and opportunities to be identified to aid decision-making [24], including “proactive/predictive models of care” [11].

Patient-Centered Care

The patient-centered care dimension encompasses the extent to which patients, caregivers, and families can actively participate in their health decisions, access information and their health data, and cocreate services and service delivery [24]. Only 3 articles in this review examined patient-centered care as a dimension of digital maturity, which resulted in 2 indicators: patient focus and patient empowerment.

The patient focus indicator assesses the extent to which the role of the patient is considered, involved, and valued when designing new models of care [25,31]. The patient empowerment indicator represents the extent to which patients are encouraged to actively participate in their health decisions and have access to relevant information and health data [24].

Discussion

Summary of Key Findings

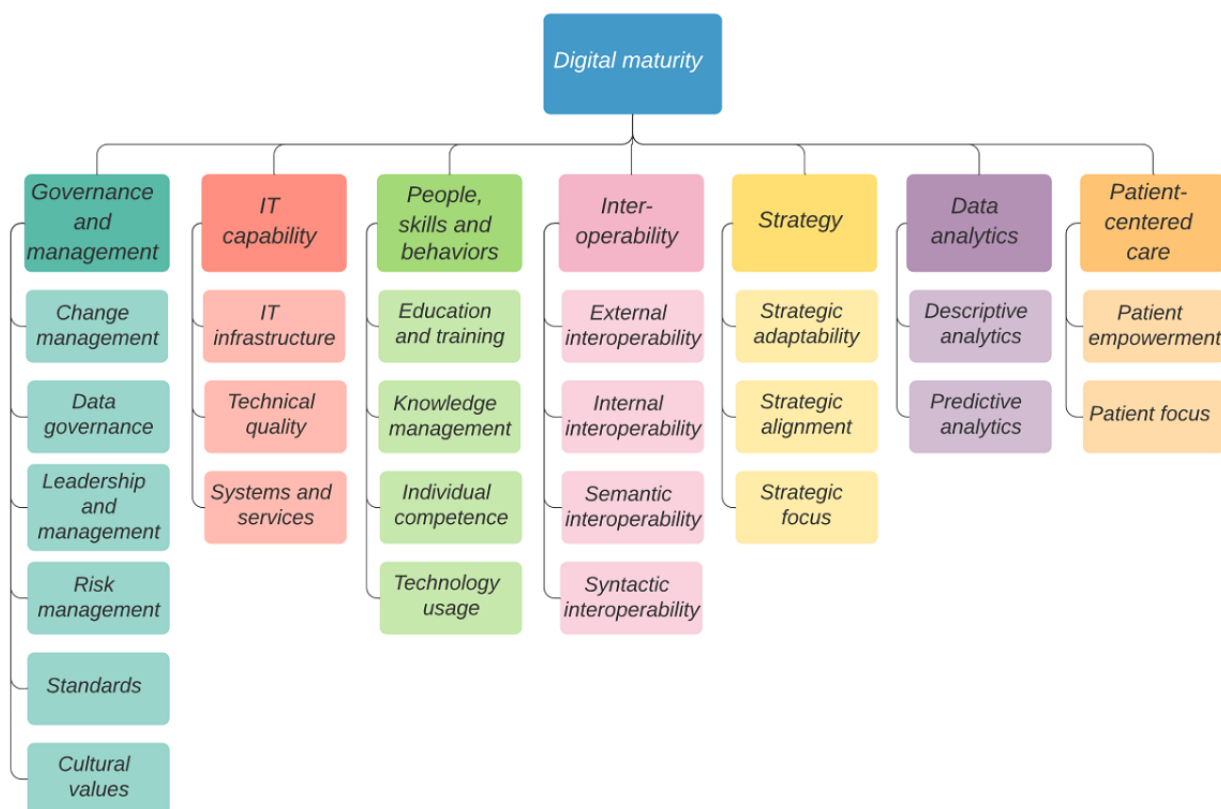
In summary, we identified 7 dimensions (ie, strategy; governance and management; IT capability; interoperability; data analytics; people, skills, and behavior; and patient-centered care) of digital health maturity that hospitals need to consider when strategically planning their digital health agenda. In addition, we identified 24 indicators that can be used to measure these dimensions (Figure 2). To operationalize these indicators, future research should seek to rigorously develop specific measurement items and follow extensive internal and external validity and reliability assessment [45].

These dimensions have received varying attention in the literature; however, as we argue in the “Implications for Future Work” section of this paper, a robust digital health maturity assessment must consider all dimensions to a sufficient depth. As such, we considered these dimensions to be equally weighted. Failure to consider a dimension could ultimately prove detrimental to the overall digital transformation agenda.

Our findings extend previous systematic literature reviews on digital health maturity models in 3 ways. First, past reviews have sought to identify various maturity models used in health

care and analyze them in isolation. For instance, Carvalho et al [37] examined 14 maturity models and Gomes and Romão [46] investigated 26 maturity models commonly employed in health care, providing a descriptive account of each. In contrast, this study synthesized the dimensions present across maturity models to derive a consolidated framework (Figure 2). Second, other reviews have investigated maturity model dimensions, yet had aggregate dimensions that were inappropriately broad. For instance, Tarhan et al [47] developed a consolidated list of only four maturity model dimensions: business process, technology, people, and other. Their business process dimension incorporated government regulations, their technology dimension was aligned with the IT capability dimension identified in this study, and their people dimension focused largely on patient safety culture and therefore differed from the people, skills, and behavior dimension identified in this study, which examined individual-level and organizational-level factors. The “other” category incorporated a wide range of factors including “culture, strategy, governance, leadership, interoperability, and data” [47]. The maturity model framework developed in this paper provides a more granular account of factors in the “other” category represented in the previously reported dimension. Such a granular account is necessary for effective assessment.

Figure 2. Consolidated Digital Maturity Model Framework for Hospitals.



Implications for Future Work

Through performing this analysis, we have identified four important areas that future research should focus on: (1) balancing digital maturity dimensions; (2) evaluating the impact

of dimensions on the quadruple aim of health care; (3) examining the interrelationships between dimensions; and (4) evaluating longitudinal variations in digital maturity. These are discussed in turn below.

Balancing Digital Maturity Dimensions

No maturity model in our review encompassed all 7 dimensions of our consolidated digital maturity model framework. The vast majority of studies focused on organizational capability (ie, governance and management and strategy), technological capability (ie, IT capability and interoperability), and individual capability (ie, people, skills, and behavior). Only 3 papers recognized patient-centered care as a dimension of digital maturity, which lags behind the goals of current medical practice. This marked difference in attention to the dimensions of the maturity models illustrates the traditional corporate focus on technical and regulatory components of digital health and the neglect of patient outcomes in the digital transformation of health care. This is a clear oversight in current digital maturity assessments, as government policies are increasingly placing the patient “at the heart of their own treatment plans so that they might develop a commitment to self-management” [48].

Moreover, while technology capability has been a prominent theme in both the IT capability and interoperability dimensions, there has been less attention paid to understanding the data analytics dimension [16]. In terms of data analytics, many of the professed benefits of digital health emanate from the “promise and potential” of the secondary use of health care data [49,50]. This capability is, further, central to prior government agendas promoting meaningful use of technology [51]. As such, future research needs to address data analytics as a key aspect of digital maturity, examining not only descriptive and predictive analytics but also the potential of prescriptive analytics.

Our findings (detailed in [Multimedia Appendix 1](#)) demonstrate that the vast majority of maturity models have been assessed in developed countries, such as the United States and the United Kingdom. Seldom is the digital maturity of hospitals in developing countries assessed (notable exceptions are the work of Yarmohammadian et al [52] and Moradi et al [31], who examined maturity models in Iran), and cross-cultural comparisons are largely overlooked. Future research therefore needs to examine the extent to which maturity models are equally applicable across cultures and settings. Ammenwerth et al [53] provide some useful guidance into how best to do so.

Evaluating the Impact of Digital Maturity Dimensions

While the importance of the 7 identified dimensions has been raised across multiple papers, their impact on outcomes such as the quadruple aim of health care (ie, reducing costs, improving patient experience, improving the work life of health care providers, and advancing population health [7]) has largely not been assessed (details are shown in [Multimedia Appendix 4](#)). Only 2 papers in our study triangulated digital maturity with outcomes. For instance, van Poelgeest et al [44] identified that the higher the digital maturity based on the Electronic Medical Record Adoption Model (EMRAM), the shorter the length of stay, although this was dependent on the location of the hospital. Conversely, Martin et al [12] identified that digital maturity based on a clinical digital maturity index did not influence the mortality, readmission, or complications encountered in the hospital, but found that maturity significantly improved length of stay and the number of harm-free patient care episodes.

Understanding how digital maturity influences outcomes is essential, as past research has found mixed results when assessing the outcomes of the digital transformation of health care, with recommendations made to policy makers to “identify and support the drivers of successful [eHealth] outcomes” [8]. If designed and applied correctly, digital maturity assessments could equip policy makers with tools to evaluate whether they have the drivers in place for successful digital transformation [54]. However, validation of the digital maturity dimensions is still required. Such validation will need to extend beyond measuring operational improvements such as cost savings and productivity goals and consider all 4 health care aims. Failure to adequately recognize the health care aim of improving the working conditions of health care providers will limit successful digital transformation, as demonstrated in many reports of staff dissatisfaction and burnout associated with digital technology in health care [55].

Similar concerns surrounding the validity of digital maturity models have been observed by Thordsen et al [56]. To validate the digital maturity of hospitals, it is necessary to analyze digital maturity through the lens of balanced health care outcomes, as outlined in the quadruple aim. We encourage future research using a multiple case study design to evaluate both the digital maturity dimensions and key performance indicators related to each aim and to assess whether digital maturity correlates with health care outcomes. Confounders will likely be present, but this is a necessary first step to provide evidence to health care executives regarding the need to evaluate and improve digital maturity. In addition, future research should seek to perform an intervention study with targeted improvements within each digital maturity dimension and assess the impact on the health care aims to further understand the mechanisms behind the purported relationship.

Understanding the Interrelationships Between Dimensions

The digital maturity dimensions in the literature reviewed here were largely examined in a subjective manner, with the dependencies and interrelationships open to interpretation, assumptions, and variability. Future research should seek to delve into these interrelationships further, as this could provide insights into the order in which hospitals should seek to improve the digital maturity dimensions. For instance, efforts to improve data analytics and IT capabilities through implementing artificial intelligence algorithms for complex clinical care may be hampered if there is no appropriate clinical governance or a clinical informatics workforce. As such, future research should seek to examine exemplar cases which have excelled in each of the dimensions to identify their drivers.

Conducting a Longitudinal Analysis of Dimensions

At different stages of a hospital’s digital health journey, different maturity model dimensions may need to be assessed. This is because digital systems and organizations are dynamic and, therefore, change over time. Although some maturity models decompose digital maturity into stages, these are often simple in nature. Some notable maturity models have taken this level of detail into account [16] by considering the varying measurement criteria between the different stages of maturity.

But as a whole this approach has mostly been overlooked. As such, scholars should seek to perform a longitudinal investigation of digital maturity to ensure appropriate assessments are performed depending on the level of IT capability within the hospital.

Limitations

This review is scoped to the digital maturity of hospitals and not to other health care settings. This is necessary because of the vast differences between acute health care settings and primary care. Future research should seek to investigate the digital maturity of primary care settings to identify maturity dimensions necessary for their successful transformation.

Although maturity models are widely being used in hospitals globally, it is important to note that digital maturity assessments are just one approach to planning and evaluating digital health transformations. Future research should compare the efficacy of digital maturity assessments with other approaches, for instance plans for digital health transformation benchmarking [53], the NASSS (nonadoption, abandonment, scale-up, spread, and sustainability) framework [57], and organizational readiness surveys [58]. Alternatively, from an evaluation perspective, organizations can adopt the measures from evaluation frameworks [59].

In addition, this literature review has been scoped to peer-reviewed outlets in medical databases and leading information systems journals, with “grey” literature excluded, which could have led to publication bias. Although this scoping may have missed some articles, it was necessary to ensure only high-quality, theoretical, rigorously developed models were included. In addition, proprietary maturity models that are used in practice but not examined in the peer-reviewed literature in this study were omitted. Nevertheless, many proprietary maturity models have been examined in peer-reviewed journals and were therefore included in this study, such as the EMRAM of the Health Information Management Systems Society.

Conclusions

This systematic literature review resulted in the development of a consolidated digital maturity model framework consisting of 7 core dimensions and 24 indicators of digital health maturity. Future research needs to be conducted to understand how these dimensions relate to outcomes across the quadruple aim of health care, and to extend the traditional IT and corporate focus to include patient and staff considerations. In that way, digital health strategic plans will become aligned to the strategic aims of hospitals and focused on delivering the quadruple aim of health care.

Acknowledgments

This publication is supported by Digital Health CRC Limited and was funded under the Commonwealth Government’s Cooperative Research Centres Programme. We acknowledge support provided by the Queensland Department of Health, the University of Queensland, the Health Care Information and Management Systems Society, and Queensland University of Technology.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Distinct Maturity Models.

[DOCX File, 30 KB - [jmir_v24i3e32994_app1.docx](#)]

Multimedia Appendix 2

Digital Maturity Dimensions and Indicators.

[DOCX File, 24 KB - [jmir_v24i3e32994_app2.docx](#)]

Multimedia Appendix 3

Mapping of Dimensions to Maturity Models.

[DOCX File, 24 KB - [jmir_v24i3e32994_app3.docx](#)]

Multimedia Appendix 4

Method Applied for Validated Maturity Model.

[DOCX File, 24 KB - [jmir_v24i3e32994_app4.docx](#)]

References

1. Johnston DS. Digital maturity: are we ready to use technology in the NHS? *Future Healthc J* 2017 Oct;4(3):189-192 [FREE Full text] [doi: [10.7861/futurehosp.4-3-189](#)] [Medline: [31098469](#)]
2. Eden R, Burton-Jones A, Casey V, Draheim M. Digital Transformation Requires Workforce Transformation. *MISQ* 2019 Mar 1;18(1):1-17. [doi: [10.17705/2msq.00005](#)]

3. Krey M. Key challenges to enabling IT in healthcare. *IEEE Int Conf Healthc Inform: IEEE*; 2016 Dec 08 Presented at: IEEE International Conference on Healthcare Informatics; October 2016; Chicago, IL, USA p. 328-333. [doi: [10.1109/ICHI.2016.60](https://doi.org/10.1109/ICHI.2016.60)]
4. Cresswell K, Sheikh A, Krasuska M, Heeney C, Franklin BD, Lane W, et al. Reconceptualising the digital maturity of health systems. *The Lancet Digital Health* 2019 Sep;1(5):e200-e201. [doi: [10.1016/s2589-7500\(19\)30083-4](https://doi.org/10.1016/s2589-7500(19)30083-4)]
5. Eden R, Burton-Jones A, Scott I, Staib A, Sullivan C. Effects of eHealth on hospital practice: synthesis of the current literature. *Aust Health Rev* 2018 Sep;42(5):568-578. [doi: [10.1071/AH17255](https://doi.org/10.1071/AH17255)] [Medline: [29986809](https://pubmed.ncbi.nlm.nih.gov/29986809/)]
6. Rahimi K. Digital health and the elusive quest for cost savings. *The Lancet Digital Health* 2019 Jul;1(3):e108-e109. [doi: [10.1016/s2589-7500\(19\)30056-1](https://doi.org/10.1016/s2589-7500(19)30056-1)]
7. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. *Ann Fam Med* 2014 Nov;12(6):573-576 [FREE Full text] [doi: [10.1370/afm.1713](https://doi.org/10.1370/afm.1713)] [Medline: [25384822](https://pubmed.ncbi.nlm.nih.gov/25384822/)]
8. Eden R, Burton-Jones A, Scott I, Staib A, Sullivan C. The impacts of eHealth upon hospital practice: synthesis of the current literature. *Deeble Institute Evidence Brief* 2017 Oct 17;16:1-7.
9. Zheng K, Abraham J, Novak LL, Reynolds TL, Gettinger A. A Survey of the Literature on Unintended Consequences Associated with Health Information Technology: 2014–2015. *Yearb Med Inform* 2018 Mar 06;25(01):13-29. [doi: [10.15265/iy-2016-036](https://doi.org/10.15265/iy-2016-036)]
10. Mettler T, Pinto R. Evolutionary paths and influencing factors towards digital maturity: An analysis of the status quo in Swiss hospitals. *Technol Forecast Soc Change* 2018 Aug;133:104-117. [doi: [10.1016/j.techfore.2018.03.009](https://doi.org/10.1016/j.techfore.2018.03.009)]
11. Krasuska M, Williams R, Sheikh A, Franklin BD, Heeney C, Lane W, et al. Technological Capabilities to Assess Digital Excellence in Hospitals in High Performing Health Care Systems: International eDelphi Exercise. *J Med Internet Res* 2020 Aug 18;22(8):e17022 [FREE Full text] [doi: [10.2196/17022](https://doi.org/10.2196/17022)] [Medline: [32808938](https://pubmed.ncbi.nlm.nih.gov/32808938/)]
12. Martin G, Clarke J, Liew F, Arora S, King D, Aylin P, et al. Evaluating the impact of organisational digital maturity on clinical outcomes in secondary care in England. *NPJ Digit Med* 2019;2:41 [FREE Full text] [doi: [10.1038/s41746-019-0118-9](https://doi.org/10.1038/s41746-019-0118-9)] [Medline: [31304387](https://pubmed.ncbi.nlm.nih.gov/31304387/)]
13. Eden R, Burton-Jones A, Grant J, Collins R, Staib A, Sullivan C. Digitising an Australian university hospital: qualitative analysis of staff-reported impacts. *Aust. Health Review* 2020;44(5):677-689. [doi: [10.1071/ah18218](https://doi.org/10.1071/ah18218)]
14. Klecun E. Transforming healthcare: policy discourses of IT and patient-centred care. *Eur J Inf Syst* 2017 Dec 19;25(1):64-76. [doi: [10.1057/ejis.2014.40](https://doi.org/10.1057/ejis.2014.40)]
15. Bassi J, Lau F. Measuring value for money: a scoping review on economic evaluation of health information systems. *J Am Med Inform Assoc* 2013 Jul;20(4):792-801 [FREE Full text] [doi: [10.1136/amiajnl-2012-001422](https://doi.org/10.1136/amiajnl-2012-001422)] [Medline: [23416247](https://pubmed.ncbi.nlm.nih.gov/23416247/)]
16. Carvalho JV, Rocha, Abreu A. Maturity Assessment Methodology for HISMM - Hospital Information System Maturity Model. *J Med Syst* 2019 Jan 07;43(2):35. [doi: [10.1007/s10916-018-1143-y](https://doi.org/10.1007/s10916-018-1143-y)] [Medline: [30613901](https://pubmed.ncbi.nlm.nih.gov/30613901/)]
17. Templier M, Paré G. A Framework for Guiding and Evaluating Literature Reviews. *CAIS* 2015;37:112-137. [doi: [10.17705/1cais.03706](https://doi.org/10.17705/1cais.03706)]
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ* 2009 Jul 21;339(jul21 1):b2535-b2535. [doi: [10.1136/bmj.b2535](https://doi.org/10.1136/bmj.b2535)]
19. Saldaña J. *The coding manual for qualitative researchers*. London: SAGE; 2021.
20. Gioia DA, Corley KG, Hamilton AL. *Seeking Qualitative Rigor in Inductive Research*. *Organ Res Methods* 2013 Jan;16(1):15-31. [doi: [10.1177/1094428112452151](https://doi.org/10.1177/1094428112452151)]
21. Vidal Carvalho J, Rocha, Abreu A. Maturity of hospital information systems: Most important influencing factors. *Health Informatics J* 2019 Sep;25(3):617-631 [FREE Full text] [doi: [10.1177/1460458217720054](https://doi.org/10.1177/1460458217720054)] [Medline: [28720012](https://pubmed.ncbi.nlm.nih.gov/28720012/)]
22. Martin G, Arora S, Shah N, King D, Darzi A. A regulatory perspective on the influence of health information technology on organisational quality and safety in England. *Health Informatics J* 2020 Jun;26(2):897-910 [FREE Full text] [doi: [10.1177/1460458219854602](https://doi.org/10.1177/1460458219854602)] [Medline: [31203707](https://pubmed.ncbi.nlm.nih.gov/31203707/)]
23. Potter I, Petersen T, D'Agostino M, Doane D, Ruiz P, Marti M, et al. The Virgin Islands National Information Systems for Health: vision, actions, and lessons learned for advancing the national public health agenda. *Rev Panam Salud Publica* 2018;42:e156 [FREE Full text] [doi: [10.26633/RPSP.2018.156](https://doi.org/10.26633/RPSP.2018.156)] [Medline: [31093184](https://pubmed.ncbi.nlm.nih.gov/31093184/)]
24. Grooten L, Borgermans L, Vrijhoef HJ. An Instrument to Measure Maturity of Integrated Care: A First Validation Study. *Int J Integr Care* 2018 Jan 25;18(1):10 [FREE Full text] [doi: [10.5334/ijic.3063](https://doi.org/10.5334/ijic.3063)] [Medline: [29588644](https://pubmed.ncbi.nlm.nih.gov/29588644/)]
25. Flott K, Callahan R, Darzi A, Mayer E. A Patient-Centered Framework for Evaluating Digital Maturity of Health Services: A Systematic Review. *J Med Internet Res* 2016 Apr 14;18(4):e75 [FREE Full text] [doi: [10.2196/jmir.5047](https://doi.org/10.2196/jmir.5047)] [Medline: [27080852](https://pubmed.ncbi.nlm.nih.gov/27080852/)]
26. Kulju S, Morrish W, King L, Bender J, Gunnar W. Patient Misidentification Events in the Veterans Health Administration: A Comprehensive Review in the Context of High-Reliability Health Care. *J Patient Saf* 2020 Sep 08:E290-E296. [doi: [10.1097/PTS.0000000000000767](https://doi.org/10.1097/PTS.0000000000000767)] [Medline: [32925569](https://pubmed.ncbi.nlm.nih.gov/32925569/)]
27. Randall KH, Slovensky D, Weech-Maldonado R, Patrician PA, Sharek PJ. Self-Reported Adherence to High Reliability Practices Among Participants in the Children's Hospitals' Solutions for Patient Safety Collaborative. *Jt Comm J Qual Patient Saf* 2019 Mar;45(3):164-169. [doi: [10.1016/j.jcjq.2018.10.001](https://doi.org/10.1016/j.jcjq.2018.10.001)] [Medline: [30471989](https://pubmed.ncbi.nlm.nih.gov/30471989/)]

28. Sullivan JL, Rivard PE, Shin MH, Rosen AK. Applying the High Reliability Health Care Maturity Model to Assess Hospital Performance: A VA Case Study. *Jt Comm J Qual Patient Saf* 2016 Sep;42(9):389-411. [doi: [10.1016/s1553-7250\(16\)42080-5](https://doi.org/10.1016/s1553-7250(16)42080-5)] [Medline: [27535456](https://pubmed.ncbi.nlm.nih.gov/27535456/)]
29. Brice S, Almond H. Health Professional Digital Capabilities Frameworks: A Scoping Review. *JMDH* 2020 Nov; Volume 13:1375-1390. [doi: [10.2147/jmdh.s269412](https://doi.org/10.2147/jmdh.s269412)]
30. Blondiau A, Mettler T, Winter R. Designing and implementing maturity models in hospitals: An experience report from 5 years of research. *Health Informatics J* 2016 Sep;22(3):758-767 [FREE Full text] [doi: [10.1177/1460458215590249](https://doi.org/10.1177/1460458215590249)] [Medline: [26142431](https://pubmed.ncbi.nlm.nih.gov/26142431/)]
31. Moradi T, Jafari M, Maleki MR, Naghdi S, Ghiasvand H. Quality Management Systems Implementation Compared With Organizational Maturity in Hospital. *Glob J Health Sci* 2015 Jul 27;8(3):174-182 [FREE Full text] [doi: [10.5539/gjhs.v8n3p174](https://doi.org/10.5539/gjhs.v8n3p174)] [Medline: [26493411](https://pubmed.ncbi.nlm.nih.gov/26493411/)]
32. de Boer JC, Adriani P, van Houwelingen JW, Geerts A. Game Maturity Model for Health Care. *Games Health J* 2016 Apr;5(2):87-91. [doi: [10.1089/g4h.2015.0091](https://doi.org/10.1089/g4h.2015.0091)] [Medline: [26859720](https://pubmed.ncbi.nlm.nih.gov/26859720/)]
33. Staggers N, Rodney M. Promoting usability in organizations with a new health usability model: implications for nursing informatics. *NI* 2012 2012:396 [FREE Full text] [Medline: [24199128](https://pubmed.ncbi.nlm.nih.gov/24199128/)]
34. Guédria W, Bouzid H, Bosh G, Naudet Y, Chen D. eHealth interoperability evaluation using a maturity model. *Stud Health Technol Inform* 2012;180:333-337. [Medline: [22874207](https://pubmed.ncbi.nlm.nih.gov/22874207/)]
35. Kose I, Rayner J, Birinci S, Ulgu MM, Yilmaz I, Guner S, HIMSS Analytics Team, et al. Adoption rates of electronic health records in Turkish Hospitals and the relation with hospital sizes. *BMC Health Serv Res* 2020 Oct 21;20(1):967 [FREE Full text] [doi: [10.1186/s12913-020-05767-5](https://doi.org/10.1186/s12913-020-05767-5)] [Medline: [33087106](https://pubmed.ncbi.nlm.nih.gov/33087106/)]
36. Van de Wetering R, Batenburg R. A PACS maturity model: a systematic meta-analytic review on maturation and evolvability of PACS in the hospital enterprise. *Int J Med Inform* 2009 Feb;78(2):127-140. [doi: [10.1016/j.ijmedinf.2008.06.010](https://doi.org/10.1016/j.ijmedinf.2008.06.010)] [Medline: [18752988](https://pubmed.ncbi.nlm.nih.gov/18752988/)]
37. Carvalho JV, Rocha, Abreu A. Maturity Models of Healthcare Information Systems and Technologies: a Literature Review. *J Med Syst* 2016 Jun;40(6):1-10. [doi: [10.1007/s10916-016-0486-5](https://doi.org/10.1007/s10916-016-0486-5)] [Medline: [27083575](https://pubmed.ncbi.nlm.nih.gov/27083575/)]
38. Williams PA, Lovelock B, Cabarrus T, Harvey M. Improving Digital Hospital Transformation: Development of an Outcomes-Based Infrastructure Maturity Assessment Framework. *JMIR Med Inform* 2019 Jan 11;7(1):e12465 [FREE Full text] [doi: [10.2196/12465](https://doi.org/10.2196/12465)] [Medline: [30632973](https://pubmed.ncbi.nlm.nih.gov/30632973/)]
39. Khanbhai M, Flott K, Darzi A, Mayer E. Evaluating Digital Maturity and Patient Acceptability of Real-Time Patient Experience Feedback Systems: Systematic Review. *J Med Internet Res* 2019 Jan 14;21(1):e9076. [doi: [10.2196/jmir.9076](https://doi.org/10.2196/jmir.9076)]
40. Kouroubali A, Papastilianou A, Katehakis DG. Preliminary Assessment of the Interoperability Maturity of Healthcare Digital Services vs Public Services of Other Sectors. *Stud Health Technol Inform* 2019 Aug 21;264:654-658. [doi: [10.3233/SHTI190304](https://doi.org/10.3233/SHTI190304)] [Medline: [31438005](https://pubmed.ncbi.nlm.nih.gov/31438005/)]
41. Orenstein EW, Muthu N, Weitkamp AO, Ferro DF, Zeidlhack MD, Slagle J, et al. Towards a Maturity Model for Clinical Decision Support Operations. *Appl Clin Inform* 2019 Oct 30;10(05):810-819. [doi: [10.1055/s-0039-1697905](https://doi.org/10.1055/s-0039-1697905)]
42. Corrigan D, McDonnell R, Zarabzadeh A, Fahey T. A Multistep Maturity Model for the Implementation of Electronic and Computable Diagnostic Clinical Prediction Rules (eCPRs). *EGEMS (Wash DC)* 2015;3(2):1153 [FREE Full text] [doi: [10.13063/2327-9214.1153](https://doi.org/10.13063/2327-9214.1153)] [Medline: [26290890](https://pubmed.ncbi.nlm.nih.gov/26290890/)]
43. Van de Wetering R, Batenburg R, Lederman R. Evolutionistic or revolutionary paths? A PACS maturity model for strategic situational planning. *Int J Comput Assist Radiol Surg* 2010 Jul;5(4):401-409. [doi: [10.1007/s11548-010-0414-y](https://doi.org/10.1007/s11548-010-0414-y)] [Medline: [20379793](https://pubmed.ncbi.nlm.nih.gov/20379793/)]
44. Van Poelgeest R, van Groningen JT, Daniels JH, Roes KC, Wiggers T, Wouters MW, et al. Level of Digitization in Dutch Hospitals and the Lengths of Stay of Patients with Colorectal Cancer. *J Med Syst* 2017 May;41(5):84 [FREE Full text] [doi: [10.1007/s10916-017-0734-3](https://doi.org/10.1007/s10916-017-0734-3)] [Medline: [28391455](https://pubmed.ncbi.nlm.nih.gov/28391455/)]
45. MacKenzie, Podsakoff, Podsakoff. Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques. *MIS Quarterly* 2011;35(2):293-334. [doi: [10.2307/23044045](https://doi.org/10.2307/23044045)]
46. Gomes J, Romão M. Information System Maturity Models in Healthcare. *J Med Syst* 2018 Oct 16;42(12):235. [doi: [10.1007/s10916-018-1097-0](https://doi.org/10.1007/s10916-018-1097-0)] [Medline: [30327955](https://pubmed.ncbi.nlm.nih.gov/30327955/)]
47. Kolukisa Tarhan A, Garousi V, Turetken O, Söylemez M, Garossi S. Maturity assessment and maturity models in health care: A multivocal literature review. *Digit Health* 2020;6:1-20 [FREE Full text] [doi: [10.1177/2055207620914772](https://doi.org/10.1177/2055207620914772)] [Medline: [32426151](https://pubmed.ncbi.nlm.nih.gov/32426151/)]
48. Barrett M, Oborn E, Orlikowski W. Creating Value in Online Communities: The Sociomaterial Configuring of Strategy, Platform, and Stakeholder Engagement. *Inf Syst Res* 2016 Dec;27(4):704-723. [doi: [10.1287/isre.2016.0648](https://doi.org/10.1287/isre.2016.0648)]
49. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;2:3 [FREE Full text] [doi: [10.1186/2047-2501-2-3](https://doi.org/10.1186/2047-2501-2-3)] [Medline: [25825667](https://pubmed.ncbi.nlm.nih.gov/25825667/)]
50. Hersh W. Editorial: Adding value to the electronic health record through secondary use of data for quality assurance, research, and surveillance. *Am J Manag Care* 2007;13(6):277-278. [doi: [10.1108/jhca.2007.06220baa.001](https://doi.org/10.1108/jhca.2007.06220baa.001)]
51. Jha AK. Meaningful use of electronic health records: the road ahead. *JAMA* 2010 Oct 20;304(15):1709-1710. [doi: [10.1001/jama.2010.1497](https://doi.org/10.1001/jama.2010.1497)] [Medline: [20959581](https://pubmed.ncbi.nlm.nih.gov/20959581/)]

52. Yarmohammadian MH, Tavakoli N, Shams A, Hatampour F. Evaluation of organizational maturity based on people capacity maturity model in medical record wards of Iranian hospitals. *J Educ Health Promot* 2014;3:18-24 [FREE Full text] [doi: [10.4103/2277-9531.134743](https://doi.org/10.4103/2277-9531.134743)] [Medline: [25077147](https://pubmed.ncbi.nlm.nih.gov/25077147/)]
53. Ammenwerth E, Duftschmid G, Al-Hamdan Z, Bawadi H, Cheung NT, Cho K, Kirca, et al. International Comparison of Six Basic eHealth Indicators Across 14 Countries: An eHealth Benchmarking Study. *Methods Inf Med* 2020 Dec;59(S 02):e46-e63 [FREE Full text] [doi: [10.1055/s-0040-1715796](https://doi.org/10.1055/s-0040-1715796)] [Medline: [33207386](https://pubmed.ncbi.nlm.nih.gov/33207386/)]
54. Burmann A, Meister S. Practical Application of Maturity Models in Healthcare: Findings from Multiple Digitalization Case Studies. *HEALTHINF* 2021:100-110. [doi: [10.5220/0010228601000110](https://doi.org/10.5220/0010228601000110)]
55. Downing NL, Bates DW, Longhurst CA. Physician Burnout in the Electronic Health Record Era: Are We Ignoring the Real Cause? *Ann Intern Med* 2018 Jul 03;169(1):50-51. [doi: [10.7326/M18-0139](https://doi.org/10.7326/M18-0139)] [Medline: [29801050](https://pubmed.ncbi.nlm.nih.gov/29801050/)]
56. Thordsen T, Murawski M, Bick M. How to Measure Digitalization? A Critical Evaluation of Digital Maturity Models. In: *Responsible Design, Implementation and Use of Information and Communication Technology*. 2020;1. Skukuza, South Africa: Springer International Publishing; 2020:358-369.
57. Greenhalgh T, Wherton J, Papoutsis 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](https://doi.org/10.2196/jmir.8775)] [Medline: [29092808](https://pubmed.ncbi.nlm.nih.gov/29092808/)]
58. Hussain M, Papastathopoulos A. Organizational readiness for digital financial innovation and financial resilience. *Int J Prod Econ* 2022 Jan;243:1-15. [doi: [10.1016/j.ijpe.2021.108326](https://doi.org/10.1016/j.ijpe.2021.108326)]
59. Burton-Jones A, Akhlaghpour S, Ayre S, Barde P, Staib A, Sullivan C. Changing the conversation on evaluating digital transformation in healthcare: Insights from an institutional analysis. *Inf Organ* 2020 Mar;30(1):1-16. [doi: [10.1016/j.infoandorg.2019.100255](https://doi.org/10.1016/j.infoandorg.2019.100255)]

Abbreviations

EMRAM: Electronic Medical Record Adoption Model

IT: information technology

MM: maturity model

NASSS: nonadoption, abandonment, scale-up, spread, and sustainability

PACS: picture archiving and communication system

Edited by A Mavragani; submitted 18.08.21; peer-reviewed by S Meister, G Martin, K Cresswell; comments to author 08.10.21; revised version received 02.12.21; accepted 28.12.21; published 30.03.22.

Please cite as:

Duncan R, Eden R, Woods L, Wong I, Sullivan C

Synthesizing Dimensions of Digital Maturity in Hospitals: Systematic Review

J Med Internet Res 2022;24(3):e32994

URL: <https://www.jmir.org/2022/3/e32994>

doi: [10.2196/32994](https://doi.org/10.2196/32994)

PMID: [35353050](https://pubmed.ncbi.nlm.nih.gov/35353050/)

©Rhona Duncan, Rebekah Eden, Leanna Woods, Ides Wong, Clair Sullivan. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 30.03.2022. 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.

Review

Physician Burnout and the Electronic Health Record Leading Up to and During the First Year of COVID-19: Systematic Review

Clemens Scott Kruse^{1*}, MSIT, MHA, MBA, PhD; Michael Mileski^{1*}, DC; Gevin Dray¹, MHA; Zakia Johnson^{1*}, MHA; Cameron Shaw^{1*}, MHA; Harsha Shirodkar¹, MHA

School of Health Administration, College of Health Professions, Texas State University, San Marcos, TX, United States

* these authors contributed equally

Corresponding Author:

Clemens Scott Kruse, MSIT, MHA, MBA, PhD

School of Health Administration

College of Health Professions

Texas State University

601 University Dr

San Marcos, TX, 78666

United States

Phone: 1 5122454462

Fax: 1 5122458712

Email: scottkruse@txstate.edu

Abstract

Background: Physician burnout was first documented in 1974, and the electronic health record (EHR) has been known to contribute to the symptoms of physician burnout. Authors pondered the extent of this effect, recognizing the increased use of telemedicine during the first year of COVID-19.

Objective: The aim of this review was to objectively analyze the literature over the last 5 years for empirical evidence of burnout incident to the EHR and to identify barriers to, facilitators to, and associated patient satisfaction with using the EHR to improve symptoms of burnout.

Methods: No human participants were involved in this review; however, 100% of participants in studies analyzed were adult physicians. We queried 4 research databases and 1 targeted journal for studies commensurate with the objective statement from January 1, 2016 through January 31, 2021 (n=25).

Results: The hours spent in documentation and workflow are responsible for the sense of loss of autonomy, lack of work-life balance, lack of control of one's schedule, cognitive fatigue, a general loss of autonomy, and poor relationships with colleagues. Researchers have identified training, local customization of templates and workflow, and the use of scribes as strategies to alleviate the administrative burden of the EHR and decrease symptoms of burnout.

Conclusions: The solutions provided in the literature only addressed 2 of the 3 factors (workflow and documentation time) but not the third factor (usability). Practitioners and administrators should focus on the former 2 factors because they are within their sphere of control. EHR vendors should focus on empirical evidence to identify and improve the usability features with the greatest impact. Researchers should design experiments to explore solutions that address all 3 factors of the EHR that contribute to burnout.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020201820; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=201820

International Registered Report Identifier (IRRID): RR2-10.2196/15490

(*J Med Internet Res* 2022;24(3):e36200) doi:[10.2196/36200](https://doi.org/10.2196/36200)

KEYWORDS

electronic health record; physician burnout; quality improvement; psychiatry; medical informatics; COVID-19; pandemic; health informatic; health care; health care professional; health care infrastructure; health care system; mental health; cognitive fatigue

Introduction

Background

This systematic review examined the state of physician burnout incident to the electronic health record (EHR), compounded by the stress of managing the pandemic in the first year of COVID-19. Neither physician burnout nor the EHR are new; however, the additional stress of managing a pandemic may make the relationship between these 2 variables clearer. The clinical psychologist Herbert Freudenberger [1] is attributed to the first mention of physician burnout in 1974, as he observed physician interaction in the drug-addled East Village of New York City. His description of burnt-out physicians mirrored the physicians' description of burnt-out patients with drug addiction in terms of a feeling of disassociation as depicted by the definition in the following sections. Physician burnout can be detrimental to physician well-being and to the quality of care provided and can result in higher turnover [2-4]. It is a significant problem that has been attributed to the EHR.

Rationale

The EHR has become a pervasive entity in the lives of all health care workers. Very few processes in the health care field are independent of the EHR. This "digital version of the patient's chart is a real-time, patient-centered record that makes information available instantly and securely to authorized users" [5]. Physician burnout is "a long-term stress reaction marked by emotional exhaustion, depersonalization, and a lack of sense of personal accomplishment" [6]. Physician burnout was already identified as a worldwide health issue before COVID-19, and digital tools such as the EHR are cited as a contributing factor to this issue [7,8]. Factors associated with the EHR cited in relation to physician burnout are usability, workflow, and documentation time [8-13]. The documentation inherent to the EHR requires significant time, as much as 2:1 hours of direct clinical face-to-face time and as much as 2 hours outside of office hours [14]. Some authors list burn-out as a new pandemic and a new normal [15,16].

A systematic review of 182 studies on a similar topic was conducted in 2018. It examined physician burnout data over a 17-year period. It identified a high incidence of physician burnout, but it failed to attribute the EHR as a contributor [17]. Another systematic review of 50 studies was conducted in 2019. It identified 4 interventions (teamwork, time management, transitions, and technology) to assuage the effects of physician burnout [10]. A systematic review in 2020 of 81 studies found interventions to decrease the digital-tool burden (training, reduced documentation and task time, expanded care teams, leveraged quality improvement and processes in workflows) in 68% of articles analyzed [9].

Objectives

The purpose of this research was to examine physician burnout issues incident to the EHR prior to and during the first year of the COVID-19 pandemic by analyzing the literature from the last 5 years. We defined physician burnout as emotional exhaustion, depersonalization, and lack of sense of personal accomplishment [6]. We examined facilitators and barriers to

the adoption of mitigation strategies of burnout incident to the EHR.

Methods

Protocol and Registration

Authors of this systematic review followed the protocol by Kruse [18] for conducting a systematic review and reported results in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) [19]. The research was registered with PROSPERO on August 31, 2020.

Eligibility Criteria

To be eligible for this study, articles had to be published in English in peer-reviewed, academic journals between January 2016 and January 2021. All study designs were accepted including both quantitative and qualitative studies with humans of all ages; however, other systematic reviews were excluded from the selection.

Information Sources

On January 29, 2021, we used a standard search string to query 4 databases: PubMed (MEDLINE), CINAHL (exclude MEDLINE), Web of Science, and Science Direct. We also performed a journal-specific search of the Mayo Clinic Proceedings.

Search Strategy

We created a Boolean search string to combine key terms listed in the Medical Subject Headings (MeSH) of the US Library of Medicine [("electronic health record" OR "electronic medical record") AND ("physician burnout") AND COVID-19]. We used the same search strategy in all databases. We used similar filter strategies in each database, because not all databases offer the same tools.

Study Selection Process

In accordance with the protocol by Kruse [18], we searched key terms in all databases, filtered results, and screened abstracts for applicability. Reviewers rejected articles if they did not produce results (were not research), such as protocols, opinions, or did not address physician burnout and use of the EHR.

Data Collection Process

We used an Excel spreadsheet as a data extraction tool, collecting additional data at each step of the process. This spreadsheet was standardized in the protocol by Kruse [18]. We used a series of 3 consensus meetings. The first consensus meeting was held after abstract screening. Subsequent consensus meetings identified observations and themes.

Data Items

In accordance with the protocol by Kruse [18], we collected the following fields of data at each step: PICOS (participants, intervention, results compared to the control group, health outcomes, study design), bias, effect size, country of origin, statistics used, strength of evidence, quality of evidence, and 3 data fields specific to the objective of this systematic review (patient satisfaction, barriers, and facilitators). Data items and

observations became the subject of the second and third consensus meetings.

Risk of Bias Assessment and Reporting

We observed bias and assessed the quality of each study using the Johns Hopkins Nursing tool for Evidence Based Practice (JHNEBP) [20]. We considered the instances of bias in how to interpret the results because bias can limit external validity.

Effect Measures

Because we accepted mixed methods and qualitative studies, we were unable to standardize summary measures as would be performed in a meta-analysis. Effect size was not reported in any study of the group for analysis.

Synthesis Methods

During the screening process, reviewers compared elements of the abstract against the objective statement of this review. Article abstracts that matched our objective statement were marked for inclusion. The rest of this subheading is for

meta-analyses—not for systematic reviews. Although the protocol by Kruse [18] for conducting a systematic review uses elements of a meta-analysis, it falls short of this standard.

Additional Analyses

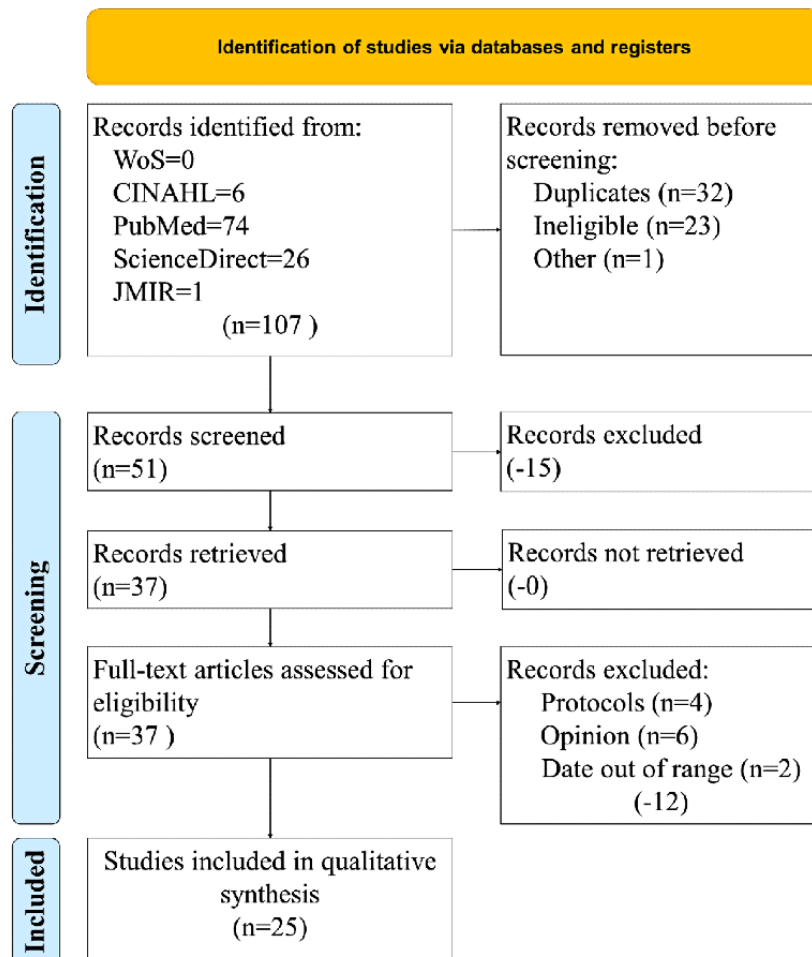
We performed a narrative analysis of the observations to convert them into themes (common threads between articles) [21]. We calculated a frequency of occurrence and reported these in a series of affinity matrices. This technique does not imply a level of importance of these observations, but instead, it simply illustrates the probability of occurrence of these observations across the group for analysis.

Results

Study Selection

Figure 1 illustrates our study selection process from the 4 databases and 1 targeted journal search. A kappa statistic was calculated based on levels of agreement between reviewers (k=0.64, moderate agreement) [22,23].

Figure 1. Study selection process. JMIR: Journal of Medical Internet Research; WoS: Web of Science.



Study Characteristics

In accordance with PRISMA 2020, a PICOS table was created from the group of articles analyzed (see Table 1). Of the 25 articles analyzed over the 5-year period, 100% of the participants were adult physicians, and all studies used the EHR as at least one of their foci in their study. Interventions ranged from using

the EHR to implementing EHR training or physician partners or scribes. Results varied across studies. Many researchers found training, education, scribes, or physician partners significantly reduced symptoms of physician burnout. Additional explanation of these results will be provided below. Interventions to reduce physician burnout noted improvements in physical pain and psychological outlook. More than half (13/25, 52%) of the study

designs were qualitative in nature. Studies are ordered as most recent to oldest: 2021 (n=2) [24,25], 2020 (n=4) [11,26-28], 2019 (n=6) [29-34], 2018 (n=8) [35-42], 2017 (n=2) [43,44], and 2016 (n=2) [45,46].

The 25 studies examined physician burnout with some intervention of the EHR before and during the COVID-19 pandemic. Of the 25 studies, 13 (52%) were qualitative studies, 4 (16%) were mixed methods, 2 sets of 2 (16%) were pre-post or observational, and 3 individual studies (12%) were

cross-sectional, cohort, or a meta-analysis. Either scribes or physician partners to enter data into the EHR during the encounter were used in 2 studies [40,46]. This intervention resulted in a decrease in symptoms of physician burnout with zero effect on patient satisfaction. EHR training or a sprint improvement process (customizing local tools) was used by 2 studies to help physicians become more efficient with the EHR [33,41]. These studies also saw a decrease in symptoms of physician burnout with zero effect on patient satisfaction.

Table 1. PICOS (participants, intervention, results [compared with a control], outcome, and study design) characteristics of the included studies.

Authors	Participants	Intervention	Result themes	Medical outcome themes	Study design
Hu et al [24]	Adult health care professionals in the ICU ^a (1122 or 46.54% doctors, 1289 or 53.46% nurses)	EHR ^b	Low frequency of exercise, comorbidities, high-quality hospital has high expectations, more night shifts, longer on the job, few paid vacations	None reported	Qualitative
Rialon et al [25]	Adult health care professionals in pediatrics (68% male, 84% White, 42-60 years old)	EHR	Long hours or workload, no time for themselves, poor work-life balance, loss of autonomy, poor relationships with colleagues	None reported	Qualitative
Giess et al [27]	Adult nonradiologists and radiologists	EHR	Radiologists more likely to report symptoms of burnout	None reported	Qualitative
Kinslow et al [28]	Adult health care professionals (41, 50.6% identified as male; 39, 48.1% identified as female; 1, 1.2% preferred not to answer; 62, 76.5% reported being a resident in a community teaching hospital; 19, 23.5% reported being a resident in a university hospital)	EHR	Women at higher risk of burnout and more likely to report suicidal ideations, poor work-life balance, long hours or workload, community-affiliated residents more likely to report suicidal ideation	None reported	Qualitative
Anderson et al [26]	Adult family medicine trainees (post-graduate years 1 through 3) and 10 family medicine faculty at the University of Arizona College of Medicine-Phoenix Family Medicine Residency	EHR	Long hours or workload	None reported	Observational
Khairat et al [11]	Adult physicians completing an EHR simulation activity, 52% female, mean age 33.2 years	EHR	Cognitive fatigue, design issues	Physical fatigue, cognitive weariness	Cross-sectional
Murphy et al [31]	Adult physicians (68% primary care physicians, 32% specialists) at 6 large health care organizations using 4 different EHR systems	EHR	Message complexity, design issues, cognitive fatigue, poor relationships with colleagues, message content	None reported	Qualitative
Tran et al [34]	Adult faculty physicians at 10 university-affiliated primary care clinics; survey sent to 190 faculty members and completed by 107 (56%) providers (86 physicians [MD/DO], 19 advanced practice providers [NP/PA], 2 providers who declined to answer the question); women = approximately two-thirds of the survey respondents; majority of the providers trained in family medicine (57%), internal medicine (27%), or pediatrics (18%)	EHR	Long hours or workload, poor work-life balance	None reported	Qualitative
Gardner et al [29]	Adult practicing physicians in Rhode Island	EHR	EHR-related or work-related stress	Work stress	Qualitative
Kroth et al [30]	Adult ambulatory primary care and subspecialty clinicians from 3 institutions (85.5% physicians, 56.7% women, 68.4% worked in primary care)	EHR	Design issues, lack of interoperability, poor work-life balance, seated position caused problems with back or wrist pain and posture	Posture, back pain	Qualitative
Sieja et al [33]	Adult clinicians in endocrinology, neurology, hematology, obstetrics, and gynecology as well as advanced practice providers	EHR Sprint process improvement	Long hours or workload	None reported	Pre-post
Quinn et al [32]	Adult physicians with an EHR	EHR	Design issues	None reported	Mixed methods
Robinson and Kersey [41]	Adult physicians from 30 specialties completing a total of 46 trainings from 2014 to 2016	EHR training	EHR improves quality and safety, readability, clinical workflow, and accuracy of documentation; efficiency gains with training; system speed and reliability issues	None reported	Mixed methods

Authors	Participants	Intervention	Result themes	Medical outcome themes	Study design
Pozdnyakova et al [40]	Adult faculty and a convenience sample (n=325) of their patients at an academic clinic (of patients: 69% Black, 65% female, 48% >65 years old); 373 patients completed surveys; 48 (13%) excluded due to incomplete data, and 325 analyzed (166 scribed and 159 nonscribed visits; Figure 1)	Scribes to assist with EHR workload	Long hours or workload	None reported	Pre-post
Marmor et al [39]	Adult physicians of internal medicine, cardiology, and gastroenterology	EHR	Time spent in EHR affects patient satisfaction	None reported	Meta-analysis
Denton et al [35]	Adult physicians at 2 urban emergency departments	EHR	EHR improves clinical workflow, door-to-doctor and time to decision, and quality and safety	None reported	Qualitative
Kroth et al [38]	Adult clinicians from 2 focus groups at 3 health care facilities with different EHRs (71% women, 98% physicians, 73% worked in primary care for an average of 11 years)	EHR	Long hours or workload, EHR-related or work-related stress, poor work-life balance	Eye strain, hand or wrist pain, back pain	Qualitative
Hauer et al [36]	Adult member and nonmember physicians practicing in Wisconsin whose email address is listed in the Wisconsin Medical Society's database	EHR	Loss of autonomy, poor relationships with colleagues, loss of autonomy, poor work-life balance	None reported	Qualitative
Young et al [42]	Adult family physician attendings, residents, and their ambulatory patients in 982 visits in clinics affiliated with 10 residencies of the Residency Research Network of Texas	EHR	Long hours or workload	None reported	Observational
Khairat et al [37]	Adult ED physicians at a large tertiary academic hospital, 50% female, 43% residents, 57% attendings	EHR	Design issues, long hours or workload, system speed or reliability issues	None reported	Mixed methods
Arndt et al [47]	Adult family medicine physicians in a single system in southern Wisconsin (100% Epic users; 43% female)	EHR	Long hours or workload	None reported	Cohort
Shahmoradi et al [44]	Adult workforce at 15 ambulatory hospitals (67% female, 75.05% with at least a BSc degree, 45.5% with age of 31-41 years, 46.67% employed <15 years)	EHR	Design issues	None reported	Qualitative
Gregory et al [43]	Adult primary care physicians at a large medical center	EHR alerts	Alert fatigue, cognitive fatigue	Physical fatigue, cognitive weariness	Mixed methods
Jamoom et al [45]	Adult physicians	EHR	Long hours or workload, longer on the job	None reported	Qualitative
Reuben et al [46]	Adult physicians were surveyed, including the pilot physicians and others who had experienced ≥1 session with a physician partner	Physician partners to help with EHR workload	Scribes or physician partners can decrease symptoms of burnout.	None reported	True experiment

^aICU: intensive care unit.

^bEHR: electronic health record.

Risk of Bias Within and Across Studies

The JHNEBP quality assessment tool was used to identify the strength and quality of evidence in the literature. These are illustrated in Table 2. Of the articles, 80% (20/25) had a strength of III, and 88% (22/25) were quality B. This means a vast majority of articles were qualitative, mixed methods, nonexperimental, or quasi-experimental in nature, but their

quality was still strong. Regarding the strength of evidence, level I studies were randomized controlled trials or true experiments. Level II studies were quasi-experimental in nature (no randomization). Level III studies were nonexperimental studies or qualitative studies. We did not accept any studies with a strength of evidence lower than III because these categories are opinion rather than research. Regarding the distribution of the 3 levels of evidence quality, in quality

category A, research shows consistent results with sufficient sample sizes, adequate controls, and definitive conclusions. In quality category B, research shows reasonably consistent results,

sufficient sample sizes, some control, and fairly definitive conclusions. As illustrated, we did not encounter any studies with a quality rating of C.

Table 2. Summary of strength and quality of evidence identified with the Johns Hopkins Nursing tool for Evidence Based Practice (JHNEBP; n=25).

Assessment	Frequency, n
Strength of evidence	
I	2
II	3
III	20
Quality of evidence	
A	3
B	22
C	0

Results of Individual Studies

Reviewers independently recorded observations for each article commensurate with the objective statement. A thematic analysis was conducted to make sense of the data. When an observation was identified more than once, it became a theme. Themes were created to summarize the observations, but they did not always exactly match the observations. These themes can be observed in [Table 3](#). Articles are sorted by most recent to oldest. [Multimedia Appendix 1](#) and [Multimedia Appendix 2](#) show the

observation-to-theme match. [Multimedia Appendix 3](#) shows additional data extracted from each study.

Reviewers conducted a thematic or narrative analysis. Part of this analysis was making sense of the data. When an observation reoccurred, it became a theme. Observations without reoccurrence were just observations. Patient satisfaction, barriers, and facilitators were explored under additional analysis. Scribes and physician partners were used in 3 studies to enter data into the EHR during an appointment, but only 2 of the studies reported on patient satisfaction outcomes.

Table 3. Summary of the analysis, sorted most recent to oldest.

Authors	Patient satisfaction themes	Barrier themes	Facilitator themes
Hu et al [24]	EHR ^a time in clinic negatively affects patient satisfaction; patient dissatisfaction negatively affects doctor-patient relationship; patient dissatisfaction negatively affects physician burnout.	Not reported	Exercise relieves symptoms of burnout; annual vacation relieves symptoms of burnout.
Rialon et al [25]	Not reported	Excessive hours spent in the EHR affect work-life balance, excessive hours spent in the EHR exacerbates symptoms of physician burnout, administrative time in the EHR takes time away from clinic and patients.	Focus on mission of care relieves symptoms of burnout.
Giess et al [27]	Not reported	EHR does not help coordinate care.	Not reported
Kinslow et al [28]	Not reported	Excessive hours spent in the EHR exacerbate symptoms of physician burnout.	Small group sessions
Anderson et al [26]	Not reported	Excessive hours spent in the EHR exacerbate symptoms of physician burnout.	Not reported
Khairat et al [11]	Not reported	EHR must undergo redesign, high number of clicks per process is inefficient.	Not reported
Murphy et al [31]	Not reported	The administrative overhead of the EHR is not conducive to efficient workflow, excessive hours spent in the EHR affect work-life balance, administrative overhead of the EHR is not conducive to efficient workflow.	Local customization (eg, templates, menus) improves efficiency, localized workflow redesign relieves symptoms of burnout.
Tran et al [34]	Not reported	Excessive hours spent in the EHR exacerbate symptoms of physician burnout.	Not reported
Gardner et al [29]	Not reported	Administrative time in the EHR takes time away from clinic and patients, excessive hours spent in the EHR affect work-life balance.	Not reported
Kroth et al [30]	Not reported	EHR must undergo redesign, excessive hours spent in the EHR exacerbate symptoms of physician burnout, high number of clicks per process is inefficient, administrative time in the EHR takes time away from clinic and patients, excessive hours spent in the EHR affect work-life balance.	Not reported
Sieja et al [33]	Not reported	Administrative overhead of the EHR is not conducive to efficient workflow.	Local customization (eg, templates, menus) improves efficiency.
Quinn et al [32]	Not reported	EHR reliability and speed, some patient information is not available due to lack of interoperability, EHR must undergo redesign.	Training increases efficiency.
Robinson and Kersey [41]	Not reported	EHR training takes time away from the clinic.	Institutional endorsement of EHR increases user acceptance of EHR, training increases efficiency.
Pozdnyakova et al [40]	Patient satisfaction not affected by scribe or physician partner in clinic during exam	Some patients do not like scribes or physician partners in the exam room, excessive hours spent in the EHR exacerbate symptoms of physician burnout.	Presence of scribe or physician partner relieves symptoms of burnout, localized workflow redesign relieves symptoms of burnout.
Marmor et al [39]	Time of day affects patient satisfaction more than time spent with patient.	Excessive hours spent in the EHR exacerbate symptoms of physician burnout.	Localized workflow redesign relieves symptoms of burnout.
Denton et al [35]	Not reported	EHR must undergo redesign, high number of clicks per process is inefficient, administrative overhead of the EHR is not conducive to efficient workflow.	EHR increases safety, decreases admission decision time, and decreases length of stay.
Kroth et al [38]	Not reported	EHR must undergo redesign, EHR reliability and speed, some patient information is not available due to lack of interoperability, administrative overhead of the EHR is not conducive to efficient workflow.	Training increases efficiency, presence of scribe or physician partner relieves symptoms of burnout.

Authors	Patient satisfaction themes	Barrier themes	Facilitator themes
Hauer et al [36]	Not reported	EHR must undergo redesign, lack of supporting practice environment, EHR creates a loss of autonomy, excessive hours spent in the EHR affects work-life balance.	Not reported
Young et al [42]	Not reported	Administrative time in the EHR takes time away from clinic and patients.	Not reported
Khairat et al [37]	Not reported	EHR must undergo redesign, EHR reliability and speed.	Not reported
Arndt et al [47]	Not reported	EHR must undergo redesign, excessive hours spent in the EHR affect work-life balance, administrative overhead of the EHR is not conducive to efficient workflow.	Not reported
Shahmoradi et al [44]	Not reported	EHR reliability and speed, excessive hours spent in the EHR exacerbate symptoms of physician burnout, some patient information is not available due to lack of interoperability, administrative overhead of the EHR is not conducive to efficient workflow, EHR investment inhibits short-term profit, EHR must undergo redesign, no standardized vocabulary.	EHR enables rapid access to information, decreases duplicate testing, increases speed of delivery of care, increases accuracy of documentation, increases safety, enables computerized analysis and interpretation of data.
Gregory et al [43]	Not reported	EHR must undergo redesign, administrative overhead of the EHR is not conducive to efficient workflow.	Not reported
Jamoom et al [45]	Not reported	Not reported	Level of physician experience with EHR increases perceived usefulness of EHR
Reuben et al [46]	Patient satisfaction not affected by scribe or physician partner in clinic during exam	Scribes or physician partners cost more money.	Presence of scribe or physician partner relieves symptoms of burnout.

^aEHR: electronic health record.

Additional Analysis

Themes and individual observations were organized into tables to reflect the probability of their occurrence in the group for analysis. These affinity matrices are shown and discussed in the following sections. In the interest of saving space, only those with the greatest number of occurrences will be discussed in detail.

Study Results

Table 4 summarizes the study results observed: 12 themes and 20 individual observations were identified by the reviewers for a total of 68 occurrences in the literature.

Of 68 occurrences, 13 (19%) identified longer hours worked and increased workload as a result of using the EHR. Researchers noted respondents to surveys worked 60-80 hours per week: The extra time was largely attributed to the EHR [25,45]. Physicians spent between 17 minutes and 217 minutes per patient in the EHR, resulting in up to 33 hours per month in the EHR after work hours: These longer hours were highly attributable to symptoms of burnout [26,34]. The nonintuitive nature of the EHR negatively impacted efficiency and contributed to the longer hours [37]. This point leads to the next item most often cited: design issues. This point occurred in 7 of 68 (10%) occurrences. Observations about design were attributed to the user interface, the long length of cut-and-paste

notes required, communication and inefficient data-sharing processes, and the requirement to memorize menu and button names [11,30-32,37,44]. The long hours spent in the EHR created a poor work-life balance [25,28,30,34,36,38]. This point occurred in 6 of 68 (9%) occurrences. Many providers felt compelled to complete administrative work in the EHR from home so that they could at least be near their families while completing their workload, but this habit created tension in the household and overall impeded attempts at work-life balance. Four themes occurred 3 times (12%): EHR improves quality and safety [35,41], a general loss of autonomy [25,36], poor relationships with colleagues [25,31,36], and cognitive fatigue [11,31,43]. The increase in quality and safety appeared in the form of greater readability of notes, increased accuracy of clinician notes, a decrease in medical errors, increased clinical efficiency, and ease of data retrieval. Loss of autonomy occurred in the literature as a general lack of control over one's schedule. Poor relationships with colleagues occurred as lack of team communication, lack of supportive practice environment, and lack of time available in the clinic to build relationships. Cognitive fatigue was only subjectively queried in 1 of the 3 studies: The other 2 were objectively measured as pupillometry and a cognitive weariness index. These themes comprised 60% of the observations. Some of these themes will appear again as either facilitators or barriers to the use of the EHR to decrease physician burnout.

Table 4. Study results affinity matrix.

Study result themes or observations	Reference(s)	Frequency, n
Long hours or workload	[25,26,28,33,34,37,38,40,42,45,47]	13
Design issues	[11,30-32,37,44]	7
Poor work-life balance	[25,28,30,34,36,38]	6
EHR ^a improves quality and safety	[35,41] ^b	3
Loss of autonomy	[25,36] ^b	3
Poor relationships with colleagues	[25,31,36]	3
Cognitive fatigue	[11,31,43]	3
EHR-related or work-related stress	[29,38]	2
Efficiency gains with training	[41] ^b	2
EHR improves clinical workflow	[35,41]	2
Longer on the job	[24,45]	2
System speed or reliability issues	[11,41]	2
EHR improves accuracy of documentation	[41]	1
EHR improves readability	[41]	1
Women more likely to report suicidal ideations	[28]	1
High-quality hospital has high expectations	[24]	1
Alert fatigue	[43]	1
Community-affiliated residents more likely to report suicidal ideations	[28]	1
Comorbidities	[24]	1
EHR improves door-to-doctor and time to decision	[35]	1
Women at a higher risk of burnout	[28]	1
Few paid vacations	[24]	1
Lack of interoperability	[30]	1
Low frequency of exercise	[24]	1
Message complexity	[31]	1
Message content	[31]	1
More night shifts	[24]	1
No time for themselves	[25]	1
Radiologists more likely to report symptoms of burnout	[27]	1
Scribes or physician partners can decrease symptoms of burnout	[46]	1
Seated position causes problems with back or wrist pain and posture	[30]	1
Time spent in EHR affects patient satisfaction	[39]	1

^aEHR: electronic health record.

^bMultiple occurrences observed in one study.

Medical Outcomes Identified With the EHR and Physician Burnout

Table 5 summarizes the medical outcomes observed: 3 themes and 4 individual observations were identified by the reviewers for a total of 10 occurrences in the literature. Of the 25 articles,

20 (80%) did not report medical outcomes. Back pain [30,38], physical fatigue [11,43], and cognitive weariness [11,43] were each mentioned 2 times out of 10 observations (60%). The other medical outcomes were eye strain, work stress, hand or wrist pain, and posture [29,30,38].

Table 5. Medical outcomes identified with the electronic health record (EHR) and physician burnout.

Medical outcome theme or observation	Reference(s)	Frequency, n
Back pain	[30,38]	2
Physical fatigue	[11,43]	2
Cognitive weariness	[11,43]	2
Eye strain	[38]	1
Work stress	[29]	1
Hand or wrist pain	[38]	1
Posture	[30]	1
None reported	[24-28,31-37,39-42,44-47]	20

Patient Satisfaction Impact of EHR

This section is not entirely logical. When we designed this study, we assumed we would find more experiments. We expected to find experiments with and without the presence of the EHR or experiments with control groups to objectively measure interventions to improve physician burnout incident to the EHR. The results of the study searches did not identify any true experiments. There were only 2 pre-post studies. The only experiments identified used training or scribes to help improve physician burnout. Table 6 identifies these as well as all mentions of patient satisfaction in the group of articles analyzed.

Although patients did not prefer a scribe in the room during an exam, their presence did not negatively affect patient satisfaction in a statistically significant manner [40,46]. Only 2 other articles mentioned patient satisfaction. One article mentioned that time in the EHR negatively affects patient satisfaction, and this negatively affects both symptoms of physician burnout and the doctor-patient relationship [24]. The other article identified the time of day the physician is in the EHR during clinic time has a greater effect on patient satisfaction than the amount of time spent with patients [39].

Table 6. Patient satisfaction impact of the electronic health record (EHR) and efforts to improve physician burnout.

Patient satisfaction theme or observation	Reference(s)	Frequency, n
Patient satisfaction not affected by scribe or physician partner in clinic during exam	[40,46]	2
EHR time in clinic negatively affects patient satisfaction	[24]	1
Time of day affects patient satisfaction more than time spent with patient	[39]	1
Patient dissatisfaction negatively affects physician burnout	[24]	1
Patient dissatisfaction negatively affects doctor-patient relationship	[24]	1
Not reported	[11,25-38,41-45,47]	21

Barriers Identified With the EHR and Physician Burnout

Table 7 summarizes the barriers incident to using the EHR to mitigate symptoms of physician burnout. The reviewers identified 8 themes and 8 individual observations, for a total of 56 occurrences in the literature; 2 articles did not identify barriers [24,45].

The theme of “EHR must undergo a redesign” occurred in 12 of 58 occurrences (21%) [11,30,32,35-38,43,44,47]. Researchers echoed their participants’ pleas to improve the design of the EHR; to reduce task repetition, screen clutter, number of clicks per task, and inefficient interfaces; improve the workflow; and reduce unnecessary searching and inefficient data entry. The inefficiencies take time away from patients and make the day longer for the provider, which impacts work-life balance. The inefficiencies lead to “excessive hours spent in the EHR, which exacerbate symptoms of physician burnout.” This theme occurred in 8 of 58 occurrences (14%) [25,26,28,30,34,39,40,44]. The administrative overhead

associated with the EHR creates inefficiencies in the standard workflow of seeing patients. This theme occurred also occurred in 8 of 58 occurrences (14%) [31,33,35,38,43,44,47]. Examples of inefficiencies were excessive data entry, illogical workflow, high number of clicks per task, and multiple screens. These inefficiencies lead to excessive hours spent in the EHR, which adversely affects work-life balance and adds to daily frustration levels. This theme occurred in 6 of 58 occurrences (11%) [25,29-31,36,47]. To add to the inefficiencies, providers noted a level of frustration at the speed and reliability issues associated with the EHR [32,37,38,44]. Participants noted communication technologies and data-sharing processes that are cumbersome and counterproductive, unpredictable system response times, and lack of hardware and infrastructure to make the EHR faster and more reliable. On the topic of administrative time in the EHR, participants noted that administrative time in the EHR takes time away from the clinic and patients. This theme occurred 4 out of 58 occurrences (7%) [25,29,30,42]. Some of the inefficiencies highlighted by providers were that some patient information is not available due to lack of interoperability. This theme occurred in 3 of 58 occurrences

(5%) [32,38,44]. This lack of availability creates data overload, which complicates data integration efforts. It often prevents linking to legacy systems, and it creates barriers with data sharing between organizations. Inefficiencies like number of clicks per process encumber efficient workflows. This theme also occurred in 3 of 58 occurrences (5%) [11,30,35]. As mentioned in the table for general results, 2 studies noted that EHR users felt a loss of autonomy [25,36]. Other observations only occurred once in the literature [27,36,40,41,44,46]. One study noted that the EHR does not coordinate care [27]. A study

that used scribes or physician partners to enter data into the EHR during the exam noted that patients do not like this practice [40]. Another study that used scribes in the exam room noted the cost to the organization for this practice [40]. A study that used training to improve provider efficiency noted this training takes time away from the clinic [41]. One study noted a lack of support by the organization for EHR tools and efficiency [36]. Another study noted that the EHR does not have a standard vocabulary [44].

Table 7. Barriers to the electronic health record (EHR) and physician burnout.

Barrier theme or observation	Reference(s)	Frequency, n
EHR must undergo redesign	[11,30,32,35-38,43,44,47] ^a	12
Excessive hours spent in the EHR exacerbate symptoms of physician burnout	[25,26,28,30,34,39,40,44]	8
The administrative overhead of the EHR is not conducive to efficient workflow	[31,33,35,38,43,44,47] ^a	8
Excessive hours spent in the EHR affect work-life balance	[25,29-31,36,47]	6
EHR reliability and speed	[32,37,38,44]	4
Administrative time in the EHR takes time away from clinic and patients	[25,29,30,42]	4
Some patient information is not available due to lack of interoperability	[32,38,44]	3
High number of clicks per process is inefficient	[11,30,35]	3
EHR creates a loss of autonomy	[25,36]	2
EHR does not help coordinate care	[27]	1
Some patients do not like scribes or physician partners in the exam room	[40]	1
EHR training takes time away from clinic	[41]	1
Scribes or physician partners cost more money	[46]	1
Lack of supporting practice environment	[36]	1
No standardized vocabulary	[44]	1
EHR investment inhibits short-term profit	[44]	1
Not reported	[24,45]	2

^aMultiple occurrences observed in one study.

Facilitators Identified With the EHR and Physician Burnout

Table 8 summarizes the facilitators incident to using the EHR to mitigate symptoms of physician burnout: 6 themes and 12 individual observations were identified by the reviewers for a total of 27 occurrences in the literature. Facilitators were not identified in 11 articles [11,26,27,29,30,34,36,37,42,43,47].

The theme of “presence of a scribe or physician partner relieves symptoms of burnout” occurred in 3 of 27 occurrences (11%) [38,40,46]. Although this practice incurs a cost to the organization, the use of either a scribe or a physician partner to enter data into the EHR during the encounter enables the physician to focus on the patient rather than negotiating the EHR, and the scribe’s time entering data into the EHR can easily be offset by a savings in provider administrative time later. This practice decreases appointment time and enables the provider to work a standard day instead of spending so much time after clinic hours catching up with the administrative side of the day’s encounters. Geriatrics practices that leveraged scribes in this

manner experienced an average of 4 minutes less per encounter for an average of 48 minutes per 4-hour session. When the administrative time after the encounter was accounted for, the savings was 88 minutes per 4-hour session. Internal medicine experienced a 2 minute per patient savings for a total of 92 minutes per 4-hour session, counting administrative time saved. Another set of studies found training would increase physician efficiency in the EHR. This theme also occurred in 3 of 27 occurrences (11%) [32,38,41]. Training decreased their frustration with the system and shortened their work day. This practice improved work-life balance and decreased symptoms of burnout. A similar theme found localized workflow redesign relieves symptoms of burnout. This theme also occurred in 3 of 27 occurrences (11%) [31,39,40]. The most common workflow redesign was preparation for encounters, which also increased patient satisfaction. Similar to training and workflow redesign, it was discovered that customized templates also increased efficiencies. This theme occurred in 2 of 27 occurrences (7%) [31,33]. This practice also increased accuracy and completeness of documentation [44], which increases quality of care. The theme “small group sessions” also occurred

in 2 of 27 occurrences (7%) [28]. This theme focused on development of young providers. This development focused on emotional and professional development. These sessions also helped establish rapport among providers. Two studies highlighted how the EHR increases safety [35,44]. The readability of orders and intelligence built into the system to alert when doses are outside of a standard range increase safety and decrease admission decision time and length of stay [35]. The other observations were only identified once [11,24,25,28-32,34-38,41,43-45,47]. One study mentioned that, although improving the EHR will help with the burden of care, it also is important to schedule regular exercise to help providers cope with the stress of care [24]. One study highlighted how a focus on the mission of care, rather than the administration of the encounter, decreases symptoms of burnout [25]. One study

highlighted the ability of the EHR to rapidly access patient data, which saves the provider time searching through a paper record [44]. Based on the conclusions of other studies, it is the process of finding this information that is key. One study highlighted the importance of provider experience (years as a provider and years in the EHR) to appreciate the usefulness of this tool [45]. Institutional endorsement of the EHR is also important [41]. This is important because it increases user acceptance of the system. A study in China found that providers who take their annual vacation tended to report fewer symptoms of burnout [24]. A study in Tehran identified the capability for the EHR to enable computerized analysis; however, this capability should be found easily rather than taking time to hunt for the feature [44]. Through training programs and customization, the EHR can increase the speed of delivery of care [44].

Table 8. Facilitators to the electronic health record (EHR) and physician burnout.

Facilitator theme or observation	Reference(s)	Frequency, n
Presence of scribe or physician partner relieves symptoms of burnout	[38,40,46]	3
Training increases efficiency	[32,38,41]	3
Localized workflow redesign relieves symptoms of burnout	[31,39,40]	3
Local customization (eg, templates, menus) improves efficiency	[31,33]	2
Small group sessions	[28] ^a	2
EHR increases safety	[35,44]	2
Exercise relieves symptoms of burnout	[24]	1
Focus on mission of care relieves symptoms of burnout	[25]	1
EHR enables rapid access to information	[44]	1
Level of physician experience with EHR increases perceived usefulness of EHR	[45]	1
Institutional endorsement of EHR increases user acceptance of EHR	[41]	1
Annual vacation relieves symptoms of burnout	[24]	1
EHR decreases admission decision time	[35]	1
EHR decreases length of stay	[35]	1
EHR decreases duplicate testing	[44]	1
EHR increases speed of delivery of care	[44]	1
EHR increases accuracy of documentation	[44]	1
EHR enables computerized analysis and interpretation of data	[44]	1
Not reported	[11,26,27,29,30,34,36,37,42,43,47]	11

^aMultiple occurrences observed in one study.

Discussion

Summary of Evidence

The preponderance of evidence supports the claim that the EHR needs an overall redesign to increase efficiency of providers. However, very few empirical studies published in the studied years could be found to measure the deficiencies. One study measured pupillometry, one measured cognitive load, and another measured cognitive weariness [11,31,43], but claims of inefficiencies were largely the result of surveys. Clearly, providers spend a great deal of time in the EHR managing the administrative necessities of the system; however, studies with

training and local customization of templates and workflow greatly improved efficiencies and decreased symptoms of burnout [33,41]. The creative use of scribes and physician partners to relieve providers of some of the real-time documentation burden showed statistically significant improvement in burnout symptoms, but they come at a price of an increased cost to employ them and a slight decrease in patient satisfaction (not statistically significant) [40,46].

From the practitioners' points of view, they wanted to know what factors remain in their sphere of influence to assuage the effects of physician burnout. Factors associated with the EHR cited in relation to physician burnout were usability, workflow,

and documentation time [8-13]. Workflow can be redesigned and customized to the user, and documentation can be performed with the use of scribes or physician partners [40,46]. The remaining factor was usability, which can only be managed in a large redesign effort. Practitioners should focus on robust and ongoing training, customization of local templates, and workflow redesign. They should weigh the economics of scribes or physician partners against the decrease in symptoms of burnout. If increasing the prevalence of symptoms of burnout increases physician turnover [2-4], certainly reducing symptoms of burnout will decrease turnover. Some best practices identified in the literature to reduce burnout were taking annual vacation [24], focusing the organization on the mission of care rather than the administration of it [25], scheduling small group sessions to help emotionally equip young providers [28], institutional endorsement of the EHR [41], and the use of regular exercise to manage stress [24]. However, these techniques do not improve the usability of the EHR, but they were identified as practices to decrease the symptoms of burnout.

Future research should empirically measure the redesign factor of usability. What aspects of usability can be improved? Are navigation issues in the EHR specific to each vendor? Are there best practices from one vendor that can be applied to other vendors without infringing upon proprietary secrets? What

mental processes in the physician workflow can be directly mapped into the menus of the EHR?

Limitations

A limitation to this review is the selection of 5 years. It was originally assumed there would be a plethora of studies on the topic of physician burnout incident to the EHR, but we found a dearth of empirical studies on the topic. There were plenty of opinion articles but very little empirical evidence. This review could have been improved by expanding the time period to 10 years, but technology advances rapidly, and reconciling the observations over a decade might have been counterproductive.

Conclusion

Although physician burnout incident to the EHR has been documented, several best practices exist to overcome 2 of the 3 factors associated with the EHR: workload and documentation time. The effect of these factors can be assuaged through workload redesign, customized templates, training, and the use of physician partners or scribes in the exam room. The third factor of usability can only be overcome through a redesign of the EHR. Practitioners should focus on the former factors, which are within their sphere of control. EHR vendors should organize empirical studies to identify targeted areas of improvement to optimize the usability of the system.

Authors' Contributions

All authors contributed equally to the preparation of this manuscript. CSK served as the lead editor and supervisor. All authors contributed to the conceptualization, methodology, study collection, study analysis, and writing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Observation-to-theme conversion for results and medical outcomes.

[DOCX File, 24 KB - [jmir_v24i3e36200_app1.docx](#)]

Multimedia Appendix 2

Observation-to-theme conversion for patient satisfaction, barriers, and facilitators.

[DOCX File, 27 KB - [jmir_v24i3e36200_app2.docx](#)]

Multimedia Appendix 3

Other observations incident to review.

[DOCX File, 26 KB - [jmir_v24i3e36200_app3.docx](#)]

References

1. Freudenberger HJ. Staff burn-out. *Journal of Social Issues* 1974;30(1):159-165. [doi: [10.1111/j.1540-4560.1974.tb00706.x](#)]
2. Halbesleben J, Rathert C. Linking physician burnout and patient outcomes: exploring the dyadic relationship between physicians and patients. *Health Care Manage Rev* 2008;33(1):29-39. [doi: [10.1097/01.HMR.0000304493.87898.72](#)] [Medline: [18091442](#)]
3. Shanafelt T, Balch CM, Bechamps G, Russell T, Dyrbye L, Satele D, et al. Burnout and medical errors among American surgeons. *Ann Surg* 2010 Jun;251(6):995-1000. [doi: [10.1097/SLA.0b013e3181bfdab3](#)] [Medline: [19934755](#)]
4. Shanafelt TD, Dyrbye LN, West CP, Sinsky CA. Potential impact of burnout on the US physician workforce. *Mayo Clin Proc* 2016 Nov;91(11):1667-1668. [doi: [10.1016/j.mayocp.2016.08.016](#)] [Medline: [27814840](#)]
5. What is an electronic health record (EHR)? The Office of the National Coordinator for Health Information Technology (ONC). URL: <https://www.healthit.gov/faq/what-electronic-health-record-ehr> [accessed 2022-02-21]

6. Physician Burnout. Agency for Healthcare Research and Quality. URL: <https://www.ahrq.gov/prevention/clinician/ahrq-works/burnout/index.html> [accessed 2022-02-21]
7. Dyrbye LN, Shanafelt TD, Sinsky CA, Cipriano PF, Bhatt J, Ommaya A, et al. Burnout among health care professionals: a call to explore and address this underrecognized threat to safe, high-quality care. *NAM Perspectives* 2017 Jul 5;7(7):1. [doi: [10.31478/201707b](https://doi.org/10.31478/201707b)]
8. Tajirian T, Stergiopoulos V, Strudwick G, Sequeira L, Sanches M, Kemp J, et al. The influence of electronic health record use on physician burnout: cross-sectional survey. *J Med Internet Res* 2020 Jul 15;22(7):e19274 [FREE Full text] [doi: [10.2196/19274](https://doi.org/10.2196/19274)] [Medline: [32673234](https://pubmed.ncbi.nlm.nih.gov/32673234/)]
9. Thomas Craig KJ, Willis VC, Gruen D, Rhee K, Jackson GP. The burden of the digital environment: a systematic review on organization-directed workplace interventions to mitigate physician burnout. *J Am Med Inform Assoc* 2021 Apr 23;28(5):985-997 [FREE Full text] [doi: [10.1093/jamia/ocaa301](https://doi.org/10.1093/jamia/ocaa301)] [Medline: [33463680](https://pubmed.ncbi.nlm.nih.gov/33463680/)]
10. DeChant PF, Acs A, Rhee KB, Boulanger TS, Snowdon JL, Tutty MA, et al. Effect of organization-directed workplace interventions on physician burnout: a systematic review. *Mayo Clin Proc Innov Qual Outcomes* 2019 Dec;3(4):384-408 [FREE Full text] [doi: [10.1016/j.mayocpiqo.2019.07.006](https://doi.org/10.1016/j.mayocpiqo.2019.07.006)] [Medline: [31993558](https://pubmed.ncbi.nlm.nih.gov/31993558/)]
11. Khairat S, Coleman C, Ottmar P, Jayachander DI, Bice T, Carson SS. Association of electronic health record use with physician fatigue and efficiency. *JAMA Netw Open* 2020 Jun 01;3(6):e207385 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.7385](https://doi.org/10.1001/jamanetworkopen.2020.7385)] [Medline: [32515799](https://pubmed.ncbi.nlm.nih.gov/32515799/)]
12. Melnick ER, Dyrbye LN, Sinsky CA, Trockel M, West CP, Nedelec L, et al. The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clin Proc* 2020 Mar;95(3):476-487 [FREE Full text] [doi: [10.1016/j.mayocp.2019.09.024](https://doi.org/10.1016/j.mayocp.2019.09.024)] [Medline: [31735343](https://pubmed.ncbi.nlm.nih.gov/31735343/)]
13. Melnick ER, Harry E, Sinsky CA, Dyrbye LN, Wang H, Trockel MT, et al. Perceived electronic health record usability as a predictor of task load and burnout among US physicians: mediation analysis. *J Med Internet Res* 2020 Dec 22;22(12):e23382 [FREE Full text] [doi: [10.2196/23382](https://doi.org/10.2196/23382)] [Medline: [33289493](https://pubmed.ncbi.nlm.nih.gov/33289493/)]
14. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med* 2016 Sep 06;165(11):753. [doi: [10.7326/m16-0961](https://doi.org/10.7326/m16-0961)]
15. Drummond D. Physician burnout: its origin, symptoms, and five main causes. *Fam Pract Manag* 2015;22(5):42-47 [FREE Full text] [Medline: [26554564](https://pubmed.ncbi.nlm.nih.gov/26554564/)]
16. McBride L. By Now, Burnout Is a Given. *The Atlantic*. 2021 Jun 30. URL: <https://www.theatlantic.com/ideas/archive/2021/06/burnout-medical-condition-pandemic/619321/> [accessed 2022-02-21]
17. Rotenstein LS, Torre M, Ramos MA, Rosales RC, Guille C, Sen S, et al. Prevalence of burnout among physicians: a systematic review. *JAMA* 2018 Sep 18;320(11):1131-1150 [FREE Full text] [doi: [10.1001/jama.2018.12777](https://doi.org/10.1001/jama.2018.12777)] [Medline: [30326495](https://pubmed.ncbi.nlm.nih.gov/30326495/)]
18. Kruse CS. Writing a systematic review for publication in a health-related degree program. *JMIR Res Protoc* 2019 Oct 14;8(10):e15490 [FREE Full text] [doi: [10.2196/15490](https://doi.org/10.2196/15490)] [Medline: [31527018](https://pubmed.ncbi.nlm.nih.gov/31527018/)]
19. Page M, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement. *J Clin Epidemiol* 2021 Jun;134:103-112. [doi: [10.1016/j.jclinepi.2021.02.003](https://doi.org/10.1016/j.jclinepi.2021.02.003)] [Medline: [33577987](https://pubmed.ncbi.nlm.nih.gov/33577987/)]
20. Newhouse R, Dearholt S, Poe S, Pugh L, White K. The Johns Hopkins nursing evidence-based practice rating scale. The Johns Hopkins Hospital. 2005. URL: http://evidencebasednurse.weebly.com/uploads/4/2/0/8/42081989/jhnedp_evidence_rating_scale.pdf [accessed 2022-02-21]
21. Todd NJ, Jones SH, Lobban FA. "Recovery" in bipolar disorder: how can service users be supported through a self-management intervention? A qualitative focus group study. *J Ment Health* 2012 Apr;21(2):114-126. [doi: [10.3109/09638237.2011.621471](https://doi.org/10.3109/09638237.2011.621471)] [Medline: [22142324](https://pubmed.ncbi.nlm.nih.gov/22142324/)]
22. Light RJ. Measures of response agreement for qualitative data: Some generalizations and alternatives. *Psychological Bulletin* 1971;76(5):365-377. [doi: [10.1037/h0031643](https://doi.org/10.1037/h0031643)]
23. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med* 2012:276-282. [doi: [10.11613/bm.2012.031](https://doi.org/10.11613/bm.2012.031)]
24. Hu Z, Wang H, Xie J, Zhang J, Li H, Liu S, et al. Burnout in ICU doctors and nurses in mainland China-A national cross-sectional study. *J Crit Care* 2021 Apr;62:265-270 [FREE Full text] [doi: [10.1016/j.jcrc.2020.12.029](https://doi.org/10.1016/j.jcrc.2020.12.029)] [Medline: [33477093](https://pubmed.ncbi.nlm.nih.gov/33477093/)]
25. Rialon K, Mueller C, Ottosen M, Weintraub AS, Coakley B, Brandt ML, et al. Drivers of distress and well-being amongst pediatric surgeons. *J Pediatr Surg* 2021 May;56(5):841-848. [doi: [10.1016/j.jpedsurg.2021.01.001](https://doi.org/10.1016/j.jpedsurg.2021.01.001)] [Medline: [33487463](https://pubmed.ncbi.nlm.nih.gov/33487463/)]
26. Anderson J, Leubner J, Brown SR. EHR overtime: an analysis of time spent after hours by family physicians. *Fam Med* 2020 Feb 7;52(2):135-137. [doi: [10.22454/fammed.2020.942762](https://doi.org/10.22454/fammed.2020.942762)]
27. Giess C, Ip IK, Gupte A, Dudley JC, Healey MJ, Boland GW, et al. RE: "Self-reported burnout: comparison of radiologists to nonradiologist peers at a large academic medical center". *Acad Radiol* 2021 Jul;28(7):1035. [doi: [10.1016/j.acra.2021.01.035](https://doi.org/10.1016/j.acra.2021.01.035)] [Medline: [33714666](https://pubmed.ncbi.nlm.nih.gov/33714666/)]
28. Kinslow K, Sutherland M, McKenney M, Elkbuli A. Reported burnout among U.S. general surgery residents: A survey of the association of program directors in surgery members. *Ann Med Surg (Lond)* 2020 Dec;60:14-19 [FREE Full text] [doi: [10.1016/j.amsu.2020.10.012](https://doi.org/10.1016/j.amsu.2020.10.012)] [Medline: [33072313](https://pubmed.ncbi.nlm.nih.gov/33072313/)]

29. Gardner R, Cooper E, Haskell J, Harris DA, Poplau S, Kroth PJ, et al. Physician stress and burnout: the impact of health information technology. *J Am Med Inform Assoc* 2019 Feb 01;26(2):106-114 [[FREE Full text](#)] [doi: [10.1093/jamia/ocy145](https://doi.org/10.1093/jamia/ocy145)] [Medline: [30517663](#)]
30. Kroth PJ, Morioka-Douglas N, Veres S, Babbott S, Poplau S, Qeadan F, et al. Association of electronic health record design and use factors with clinician stress and burnout. *JAMA Netw Open* 2019 Aug 02;2(8):e199609 [[FREE Full text](#)] [doi: [10.1001/jamanetworkopen.2019.9609](https://doi.org/10.1001/jamanetworkopen.2019.9609)] [Medline: [31418810](#)]
31. Murphy DR, Satterly T, Giardina TD, Sittig DF, Singh H. Practicing clinicians' recommendations to reduce burden from the electronic health record inbox: a mixed-methods study. *J Gen Intern Med* 2019 Sep;34(9):1825-1832 [[FREE Full text](#)] [doi: [10.1007/s11606-019-05112-5](https://doi.org/10.1007/s11606-019-05112-5)] [Medline: [31292905](#)]
32. Quinn M, Forman J, Harrod M, Winter S, Fowler KE, Krein SL, et al. Electronic health records, communication, and data sharing: challenges and opportunities for improving the diagnostic process. *Diagnosis (Berl)* 2019 Aug 27;6(3):241-248 [[FREE Full text](#)] [doi: [10.1515/dx-2018-0036](https://doi.org/10.1515/dx-2018-0036)] [Medline: [30485175](#)]
33. Sieja A, Markley K, Pell J, Gonzalez C, Redig B, Kneeland P, et al. Optimization sprints: improving clinician satisfaction and teamwork by rapidly reducing electronic health record burden. *Mayo Clin Proc* 2019 May;94(5):793-802 [[FREE Full text](#)] [doi: [10.1016/j.mayocp.2018.08.036](https://doi.org/10.1016/j.mayocp.2018.08.036)] [Medline: [30824281](#)]
34. Tran B, Lenhart A, Ross R, Dorr DA. Burnout and EHR use among academic primary care physicians with varied clinical workloads. *AMIA Jt Summits Transl Sci Proc* 2019;2019:136-144 [[FREE Full text](#)] [Medline: [31258965](#)]
35. Denton CA, Soni HC, Kannampallil TG, Serrichio A, Shapiro JS, Traub SJ, et al. Emergency physicians' perceived influence of EHR use on clinical workflow and performance metrics. *Appl Clin Inform* 2018 Jul;9(3):725-733 [[FREE Full text](#)] [doi: [10.1055/s-0038-1668553](https://doi.org/10.1055/s-0038-1668553)] [Medline: [30208497](#)]
36. Hauer A, Waukau HJ, Welch P. Physician burnout in Wisconsin: an alarming trend affecting physician wellness. *WMJ* 2018 Dec;117(5):194-200 [[FREE Full text](#)] [Medline: [30674095](#)]
37. Khairat S, Burke G, Archambault H, Schwartz T, Larson J, Ratwani R. Perceived burden of EHRs on physicians at different stages of their career. *Appl Clin Inform* 2018 Apr;9(2):336-347 [[FREE Full text](#)] [doi: [10.1055/s-0038-1648222](https://doi.org/10.1055/s-0038-1648222)] [Medline: [29768634](#)]
38. Kroth P, Morioka-Douglas N, Veres S, Pollock K, Babbott S, Poplau S, et al. The electronic elephant in the room: Physicians and the electronic health record. *JAMIA Open* 2018 Jul;1(1):49-56 [[FREE Full text](#)] [doi: [10.1093/jamiaopen/ooy016](https://doi.org/10.1093/jamiaopen/ooy016)] [Medline: [31093606](#)]
39. Marmor R, Clay B, Millen M, Savides T, Longhurst C. The impact of physician EHR usage on patient satisfaction. *Appl Clin Inform* 2018 Jan;9(1):11-14 [[FREE Full text](#)] [doi: [10.1055/s-0037-1620263](https://doi.org/10.1055/s-0037-1620263)] [Medline: [29298451](#)]
40. Pozdnyakova A, Laiteerapong N, Volerman A, Feld LD, Wan W, Burnet DL, et al. Impact of medical scribes on physician and patient satisfaction in primary care. *J Gen Intern Med* 2018 Jul;33(7):1109-1115 [[FREE Full text](#)] [doi: [10.1007/s11606-018-4434-6](https://doi.org/10.1007/s11606-018-4434-6)] [Medline: [29700790](#)]
41. Robinson KE, Kersey JA. Novel electronic health record (EHR) education intervention in large healthcare organization improves quality, efficiency, time, and impact on burnout. *Medicine (Baltimore)* 2018 Sep;97(38):e12319 [[FREE Full text](#)] [doi: [10.1097/MD.00000000000012319](https://doi.org/10.1097/MD.00000000000012319)] [Medline: [30235684](#)]
42. Young RA, Burge SK, Kumar KA, Wilson JM, Ortiz DF. A time-motion study of primary care physicians' work in the electronic health record era. *Fam Med* 2018 Feb 2;50(2):91-99. [doi: [10.22454/fammed.2018.184803](https://doi.org/10.22454/fammed.2018.184803)]
43. Gregory M, Russo E, Singh H. Electronic health record alert-related workload as a predictor of burnout in primary care providers. *Appl Clin Inform* 2017 Dec 20;08(03):686-697. [doi: [10.4338/aci-2017-01-ra-0003](https://doi.org/10.4338/aci-2017-01-ra-0003)]
44. Shahmoradi L, Darrudi A, Arji G, Farzaneh Nejad A. Electronic health record implementation: a SWOT analysis. *Acta Med Iran* 2017 Oct;55(10):642-649 [[FREE Full text](#)] [Medline: [29228530](#)]
45. Jamoom EW, Heisey-Grove D, Yang N, Scanlon P. Physician opinions about EHR use by EHR experience and by whether the practice had optimized its EHR use. *J Health Med Inform* 2016 Jul 30;7(4):1000240 [[FREE Full text](#)] [doi: [10.4172/2157-7420.1000240](https://doi.org/10.4172/2157-7420.1000240)] [Medline: [27800279](#)]
46. Reuben DB, Miller N, Glazier E, Koretz BK. Frontline account: physician partners: an antidote to the electronic health record. *J Gen Intern Med* 2016 Aug;31(8):961-963 [[FREE Full text](#)] [doi: [10.1007/s11606-016-3727-x](https://doi.org/10.1007/s11606-016-3727-x)] [Medline: [27206541](#)]
47. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W, Sinsky CA, et al. Tethered to the EHR: primary care physician workload assessment using EHR event log data and time-motion observations. *Ann Fam Med* 2017 Sep;15(5):419-426 [[FREE Full text](#)] [doi: [10.1370/afm.2121](https://doi.org/10.1370/afm.2121)] [Medline: [28893811](#)]

Abbreviations

EHR: electronic health record

JHNEBP: Johns Hopkins Nursing tool for Evidence Based Practice

PICOS: participants, intervention, comparison of results to control, outcome (medical), study design

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta Analyses

Edited by C Lovis, G Eysenbach; submitted 05.01.22; peer-reviewed by J Walsh; comments to author 28.01.22; revised version received 28.01.22; accepted 30.01.22; published 31.03.22.

Please cite as:

Kruse CS, Mileski M, Dray G, Johnson Z, Shaw C, Shirodkar H

Physician Burnout and the Electronic Health Record Leading Up to and During the First Year of COVID-19: Systematic Review
J Med Internet Res 2022;24(3):e36200

URL: <https://www.jmir.org/2022/3/e36200>

doi: [10.2196/36200](https://doi.org/10.2196/36200)

PMID: [35120019](https://pubmed.ncbi.nlm.nih.gov/35120019/)

©Clemens Scott Kruse, Michael Mileski, Gevin Dray, Zakia Johnson, Cameron Shaw, Harsha Shirodkar. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 31.03.2022. 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.

Viewpoint

Patient Influencers: The Next Frontier in Direct-to-Consumer Pharmaceutical Marketing

Erin Willis^{1*}, PhD, MPH; Marjorie Delbaere^{2*}, PhD

¹Department of Advertising, Public Relations, and Media Design, University of Colorado Boulder, Boulder, CO, United States

²Edwards School of Business, University of Saskatchewan, Saskatoon, SK, Canada

* all authors contributed equally

Corresponding Author:

Erin Willis, PhD, MPH

Department of Advertising, Public Relations, and Media Design

University of Colorado Boulder

478 UCB

1511 University Ave

Boulder, CO, 80310

United States

Phone: 1 3034927161

Email: erin.willis@colorado.edu

Abstract

Social media influencers are becoming an increasingly popular strategic communication tactic used across industry verticals, including entertainment, fashion, and beauty, to engage directly with consumers. Pharmaceutical companies have also recently entered the social media marketing arena and—within the bounds of governmental regulations—have found ways to build relationships directly with patients using covert persuasion tactics like partnering with social media influencers. Due to consumers' negative perceptions of pharmaceutical companies, it makes sense that new marketing tactics are being used to establish and improve relationships with consumers. Previous research well documents the ethical dilemmas of direct-to-consumer advertising, and there is recent burgeoning literature on online covert marketing tactics. The academic and medical literature, however, is behind in regard to social media influencers used in health and medicine. This paper highlights and defines terms used in industry practice, and also calls for more investigation and sets forward a research agenda. As consumers spend more time online and patients continue to consult social media for health information, it is important that this new marketing trend does not go unnoticed.

(*J Med Internet Res* 2022;24(3):e29422) doi:[10.2196/29422](https://doi.org/10.2196/29422)

KEYWORDS

social media; influencers; health; pharmaceutical marketing; direct-to-consumer advertising; relationship marketing; marketing; advertising; pharmaceuticals; ethics

Background

Tasked with research and development, pharmaceutical companies aid in the worldwide prevention and treatment of illness and disease; these companies encounter fierce competition and (usually) work within the bounds of government regulations [1]. Pharmaceutical companies set aside a portion of their budgets for research and development but over time the amount spent on direct-to-consumer marketing has surpassed research and development [2]. Direct-to-consumer pharmaceutical marketing refers to the promotion of prescription medications to consumers as patients, instead of targeting only doctors [3]. Previous research purports that direct-to-consumer marketing is effective [4,5], especially in encouraging consumers

to ask their doctors about specific medications. A survey by the Kaiser Family Foundation shows that after talking to their physician about a medicine they saw advertised, about 44% of patients who requested the advertised medication were ultimately prescribed this medication [6]. However, patient trust in the pharmaceutical industry is extremely low—only 58% of Americans trust pharmaceutical companies [7]. This poses a challenge for pharmaceutical companies and their direct-to-consumer marketing efforts.

Health care marketers are beginning to use the term patient influencer to refer to those who promote pharmaceutical medications and/or medical devices, allowing companies to “leverage the patient experience and expertise in the design, development and promotion of their products and services” [8].

Due to consumers' negative perceptions of pharmaceutical companies, it makes sense that new tactics are being used to establish and improve relationships with consumers. A recent report by eMarketer noted that consumer response is highest when messages are delivered from social media influencers compared to brand-owned channels; further, content from influencers is more effective at meeting communication goals [9]. Pharmaceutical marketers' entry into social media in general, and influencer marketing in particular, presents both opportunities and challenges relevant to the various stakeholders involved. Little published research is available in this area, despite the pharmaceutical industry's increasing use of patient influencers. There are different forms of patient influence, and each involves patients in a different way. The purpose of this viewpoint paper then is to simply raise awareness of influencer

marketing by pharmaceutical companies and generate research and debate on the use of influencers in health care.

The following sections of this paper will explore the concepts of expert patients, patient advocates, and digital opinion leaders. Expert patients are far removed from pharmaceutical companies; they exert influence on other patients via online health communities [10,11]. A patient advocate is also active online, but they influence other patients via their social media presence and raise awareness about a particular disease or illness [12,13]. Some patient advocates are directly involved in promoting pharmaceutical products. Some health care professionals are also digital opinion leaders whose influence is due to their active social media presence along with a reputation for being a leader in their field [14,15]. Table 1 summarizes these key concepts. Together, these different groups hold the potential to involve and influence patients in health and medical decision-making.

Table 1. Key concepts.

Concept	Explanation
Expert patient	Patients who are actively involved in online communities and who share disease experiences, information, and support with other patients and develop expertise in disease self-management.
Patient advocate	Third-party users of social media who are vocal in raising awareness of illness and disease. They are a specialized type of social media influencer and are recruited by pharmaceutical companies to participate in the development and promotion of pharmaceutical products.
Digital opinion leader	Health care professionals with an active social media presence who are seen as leaders among their peers.

Online Communities and Expert Patients

Shared decision-making in health care requires collaboration between patients and physicians; in reality, however, there are many obstacles to patients collaborating with physicians, including technology barriers, health literacy levels, and access to health insurance [16]. Physicians' time is also constrained; a recent survey reported that 51% of physicians spend 9-16 minutes with patients [17], which could have an impact on health outcomes [18-20]. Patients then often turn to the internet and social media to fill the gaps in the health care system.

Online communities can be a source of health information and peer empowerment for patients. Online communities are social platforms where "people come together to get and give information or support, or learn or to find company" [21]. A study of an online community for patients with arthritis found that members developed expertise in disease self-management in part by modeling behaviors that were discussed in the community [22]. By sharing disease experience, patients are able to "crowdsource" answers to their current situation, and "see" the outcome of different strategies through the shared user-generated content [23]. A study by Fox, Ward, and O'Rourke [24] found evidence to suggest that online communities result in the development of *expert patients*. The researchers found that active participants ultimately became expert patients in the online community; they shared information within the community and received valuable support from other patients. Although these patients became experts, their expertise was limited to the biomedical model that supports pharmaceutical interventions. They did not become experts in a wide range of approaches to weight loss [24]. Despite this

limitation, the authors suggest that informed patients are in turn informed consumers who are engaging directly with providers on health technologies as opposed to only indirectly through a medical professional.

Research found that patients often join online communities in search of information about disease treatment options, including pharmaceutical medications [25-27]. For instance, a recent content analysis (N=1960) of 4 online health communities found that medication (N=568) is one of the most popular topics discussed by members [28]; medication adherence is discussed in terms of the presence of disease symptoms, and members share their experiences with specific medications, focusing on the specific barriers to adherence (eg, cost, side effects) [29]. The number of online health communities has rapidly increased as more patients desire to access alternate sources of information as well as connect with other patients with the same illness or disease [30,31]. Although previous research notes that online health education programs about chronic disease self-management positively impact participants' health outcomes [32,33], scant research studies the health effects of online peer-to-peer communities and the influence of user-generated content on health attitudes and behaviors.

Social Media Influencers

Recent marketing efforts by pharmaceutical companies have focused on digital advertising and engagement tactics to connect with patients and build relationships [34]. In fact, pharmaceutical companies are using digital data analytics to formulate target audience profiles and direct-to-consumer marketing strategies that take a holistic approach and consider patients' overall lifestyle preferences and their well-being, not only the disease

diagnosis [35]. This new focus describes pharmaceutical companies' efforts in relationship marketing, using social platforms to speak with patients and learn about their preferences [36]. Relationship marketing draws from the practice of public relations and attempts to foster customer loyalty and long-term retention.

The pop culture term *influencer* refers to a brand's commercialization of the relationship between an influential social media user and his or her followers [37]. Social media influencers can be defined as "third-party users of social media who have achieved micro-celebrity status in the form of large followings on social media platforms and who have a position of influence on their audience" [38]. From a strategic communication perspective, social media influencers are actors who influence "organizational stakeholders through content production, content distribution, interaction, and personal appearance on the social web" [39]. The relationship between an influencer and his or her followers is fundamental to a brand's success because it should drive positive word-of-mouth and purchase intentions [40]. eMarketer reported that more than two-thirds of North American retailers use some form of influencer marketing [41]. Brands are not just selling products but entire lifestyles, employing social media influencers to build trust through authentic, curated content. Followers perceive influencers as being knowledgeable experts on specific topics, with relevant sources for information and support for their opinions and behaviors [42].

Influencers may or may not have demographics and psychographics that are similar to those of the consumer, but they are similar to the consumer in that they share a common interest in the topic of the group [43]. Influencers are chosen to represent a brand for a number of reasons, including industry popularity [44], quality of social media content [45], experience with brand partnerships [46], and audience demographics [47]. Choosing an influencer is part of a larger marketing strategy that usually supports campaign goals and broader communication objectives [48].

Pharmaceutical Influencers: Patients, Advocates, and Opinion Leaders

In direct-to-consumer pharmaceutical marketing, it can be risky for brands to activate paid social media influencers for a number of reasons: navigating Food and Drug Administration (FDA) and Federal Trade Commission (FTC) regulations, concerns about authenticity, and managing consumer engagement [49-52]. Kim Kardashian's endorsement of Diclegis, a medication to treat morning sickness, is a prime example of the risks inherent in celebrity influencers partnering with pharmaceutical companies. In 2015, Kardashian and Duchesnay, the drug manufacturer, were found to have violated FDA regulations by not properly disclosing the risks and side effects of the drug in Kardashian's Instagram post about how the drug helped her combat morning sickness during her pregnancy [53,54]. Rather than continuing to pursue partnerships with celebrity influencers, pharmaceutical marketers have instead turned to health and medical opinion leaders as well as *patient advocates*. Typically, patient advocates are active on social media, tend to raise

awareness of illness and disease, and are considered micro- or nano-influencers. These types of influencers typically have a smaller number of followers but cultivate more targeted communities, generating higher engagement rates and building stronger relationships with stakeholders than celebrity influencers [55]. Pharmaceutical brands are beginning to opt for micro- and nano-influencers who have very specific audiences that may be primed for health messaging. A recent article in Vox described the value of patients to pharmaceutical marketers in building brands [56]. Lived experience is something that patients have in common with each other, and cannot be replicated; thus, patient influencers are simply commercializing their lived health experience. Patient influencers used in this way attempt to create an emotional linkage with followers by sharing strategic and curated pieces of their illness and disease experience.

According to a recent comprehensive review of medical marketing expenditures in the United States, pharmaceutical companies spend nearly 70% of their promotional budgets marketing to health care professionals, or just over US \$20 billion in 2016 [57]. These marketing activities include prescriber detailing, free samples, direct physician payments, and disease education. A substantial portion of the direct payments goes toward sponsoring key opinion leaders [58]. Physicians have long been influencing other physicians by giving keynote speeches and lectures at educational events and serving as product champions or product endorsers for pharmaceutical companies [59]. The industry now also recruits physicians who are digital opinion leaders; that is, health care professionals with an active social media presence who are seen as leaders among their peers, but who might have different characteristics than traditional key opinion leaders [58]. Pharmaceutical marketers are beginning to use the term *influencer* to describe this type of opinion leader. Online marketing services like Klear compile lists of the top physician influencers across social media platforms [60].

Blurred Lines: Patient Empowerment or Patient Deception?

Noticeably absent from the literature is the role advertising and marketing agencies play in relationship marketing and the use of influencers to promote pharmaceutical brands. Patient advocates self-select to engage in pharmaceutical marketing efforts and are supplied with promotional collateral to share with their followers so that their persuasive messaging stays within the boundaries of disease awareness—the type of direct-to-consumer advertising permitted by regulating government agencies. However, little is published about this phenomenon. Sites like WEGO Health are actively recruiting patients as social media influencers or patient advocates so as to connect pharmaceutical companies and other health and medical brands with patients to commercialize their lived disease experience [8].

Research Agenda and Conclusion

Overview

No known academic research is published on pharmaceutical influencer marketing. There are some industry papers on the topic [61-63]. Although industry practice moves quickly, most details of the relationship between pharmaceutical companies and advertising agencies or digital marketing companies are protected by intellectual property laws and nondisclosure agreements. Our purpose here is to suggest a research agenda. We argue that it is a distinct possibility that pharmaceutical companies could look next to online patient communities. What would happen if they partnered with online communities already successful at encouraging patient expertise? Although marketers might argue that this type of direct-to-consumer marketing is beneficial to creating empowered and informed patients, more research needs to be done to better understand this increasingly popular industry practice. Investigation needs to be conducted in the areas of patient influencers and patient advocates on social media and how these opinion leaders are being used in health care marketing.

To advance knowledge in this area, we propose the following research questions for investigation and intellectual debate.

Research Question 1: Are the Ethical Issues the Same for Influencer Marketing as Traditional Direct-to-Consumer Advertising Channels?

The debate on the risks and benefits of pharmaceutical direct-to-consumer advertising is complex; it is worth noting, however, that various studies of online drug advertisements have shown that in some cases suspect claims are made and in general the ads tend to overemphasize the benefits of the drug [64,65]. Direct-to-consumer advertising may also create a preference among consumers for recently launched pharmaceutical drugs over more established treatments because of heavy advertising campaigns; this can lead the market to concentrate only on newer products, excluding older, traditional options [66,67].

When pharmaceutical companies recruit patient advocates as influencers, they are in essence adapting the opinion leader strategy that has proved very successful for them with physicians to a new target market, much the same way they did with direct-to-consumer advertising. Similar concerns could also be raised with the practice of targeting consumers with a new type of marketing strategy as opposed to targeting physicians. Physicians possess many years of medical education and training that afford them expert knowledge when they are the target of persuasion tactics. They can draw on their high levels of topic knowledge in addition to their knowledge of persuasion when coping with persuasion attempts (ie, the attempts of another physician to influence their attitudes about a pharmaceutical brand). Consumers, on the other hand, do not have this same level of medical expertise and instead must rely on their knowledge of persuasion when assessing the validity of a persuasion agent's claims [68].

Research Question 2: How Does the Influence of Patient Experts, Influencers, Celebrity Patients, and

Small Patient Forums Differ From More Traditional Advertising Venues?

Future research should explore whether the new modalities enhance the credibility of messaging within a larger societal context characterized by celebrity culture, mistrust of formal institutions, and the more recent and profound social polarizations.

Influencer marketing shares similar characteristics to native advertising, including promotion that does not hint at advertising as well as not using highly persuasive messaging [69-71]. The interests of consumers are the primary consideration in native advertising, not only the brand; thus, content is usually relevant and useful to audiences [72]. The blurring of boundaries between paid and earned content and the prevalence of covert persuasion attempts raises questions of transparency, ethics, and trust, which has been the focus of native advertising research [73,74]. Online communities and social networks encourage the exchange of health care information, including symptoms, treatment options, and potential side effects, and opinions about experiences with doctors and other health care providers [75]. This online exchange of health care information has the potential to influence health care decision-making [76]. The potential impact of covert marketing tactics on health care decision-making is not yet known.

Beyond marketing strategy, we need to understand how this type of influence affects medical decision-making and what effect that has on patients' health outcomes. To what extent do influencers affect decision-making by providers and patients? To what extent have these decisions caused harm (or good)? What factors affect whether the consumer will follow the influencer or their shared recommendations? As the media environment continues to fracture and consumers' trust in government offices, pharmaceutical companies, and health care professionals declines, employing patient influencers may be a way to rebuild that trust while also improving consumers' attitudes and behaviors related to health [77-79]. Research shows that social media influencers have a high return on investment [80], but what does that mean for pharmaceutical companies and patients' health outcomes?

Research Question 3: Who Are the Patient Influencers? Who Is Recruited by Pharmaceutical Companies and Who Approaches Pharmaceutical Companies With Intent of Monetizing a Medical Condition? Who Receives Compensation, and Who Does Not?

As mentioned previously, micro- or nano-celebrities are beginning to act as patient influencers for pharmaceutical marketers. In addition to this type of celebrity, patient advocates are also playing the role of patient influencer. Future research should explore whether there are differences in the conditions surrounding the use of each type of patient influencer and whether there are differences in type or level of compensation. Additionally, future studies should explore the motivation of these different patients to act as influencers; specifically, whether some could be considered market mavens as opposed to opinion leaders. A market maven is someone with generalized market knowledge, across multiple product categories, who

enjoys helping other consumers make better marketplace decisions [81]. The concept of a market maven is related to, but distinct from, that of an opinion leader. A study that investigated an online community of ecstasy users suggests that this last characteristic is a key differentiator between market mavens and opinion leaders [81]. Opinion leaders' interactions with others are motivated by self-involvement, while market mavens' interactions with others are motivated by a desire to help others. According to O'Sullivan [82], market mavens are individuals who share marketplace information with the goal of reducing consumption-related risks of other consumers, essentially filling information gaps that exist within a market. It is our contention that the concept of a marketplace can be extended to the domain of chronic disease, in which people are consuming medications, health care aids, and perhaps most importantly, information related to their disease. We believe that the concept of market maven holds great potential for improving our understanding of the role that online health communities play in consumers' health decisions.

Research Question 4: What Is the Potential for Misinformation in Influencer Marketing? Who Is Responsible for Harm Caused in These Forums of Misinformation?

For decades, pharmaceutical companies have modified direct-to-consumer marketing strategies, being early adopters of communication platforms [83]. Burgeoning research explores pharmaceutical companies' use of branded and unbranded content on social media, including influencer marketing campaigns [84,85]. In November 2019, the FTC updated their social media marketing guidelines on how brands and influencers should work together and what is necessary to disclose [51]. However, the FTC's guidelines are vague and up

to interpretation if no pharmaceutical brand name is mentioned. So far, these campaigns appear to be limited to partnering with existing influencers to promote products.

A recent study of the general population found that social influencers impacted diet-related decisions for 32% of survey respondents [86]. The authors of the study expressed concern that in spite of this impact, most social influencers have no official qualifications as dietitians or nutritionists and frequently share information without any scientific evidence [86]. However, many of these influencers share from their own personal experience and that brings a certain type of value all its own to patients. Although this type of influence is frustrating to health care professionals, it is fascinating to marketers. People are persuaded by powerful personal stories in the absence of "hard" evidence. As more consumers turn to the internet as a primary source for health information, consideration needs to be given to how this information affects consumers' decision-making. Studies have shown that consumers can be skeptical about their doctor's motives for prescribing brand name drugs when the brand's logo appears on objects in the doctor's office [87]. Preliminary work investigating consumer skepticism with online forums managed by pharmaceutical brands suggests that consumers are less skeptical about online communication from pharmaceutical companies and do not recognize these sponsored communities as persuasion attempts [88].

As direct-to-consumer advertising continues to evolve and innovate, the use of influencer marketing will also continue to increase; this phenomenon needs more investigation as consumers spend an increasing amount of time online and engaging with social media, and as pharmaceutical companies collect more information and connect directly with patients.

Acknowledgments

This research was supported by a grant from the Social Sciences and Humanities Research Council of Canada. Publication of this article was funded by the University of Colorado Boulder Libraries Open Access Fund.

Conflicts of Interest

None declared.

References

1. Fisher JA, Cottingham MD, Kalbaugh CA. Peering into the pharmaceutical "pipeline": investigational drugs, clinical trials, and industry priorities. *Soc Sci Med* 2015 Apr;131:322-330 [FREE Full text] [doi: [10.1016/j.socscimed.2014.08.023](https://doi.org/10.1016/j.socscimed.2014.08.023)] [Medline: [25159693](https://pubmed.ncbi.nlm.nih.gov/25159693/)]
2. DiMasi JA, Grabowski HG, Hansen RW. *J Health Econ* 2016 May;47:20-33. [doi: [10.1016/j.jhealeco.2016.01.012](https://doi.org/10.1016/j.jhealeco.2016.01.012)] [Medline: [26928437](https://pubmed.ncbi.nlm.nih.gov/26928437/)]
3. Williams JR, Hensel PJ. Direct-to-consumer advertising of prescription drugs. *J Health Care Mark* 1995;15(1):35-41. [Medline: [10142385](https://pubmed.ncbi.nlm.nih.gov/10142385/)]
4. Dieringer NJ, Kukkamma L, Somes GW, Shorr RI. Self-reported responsiveness to direct-to-consumer drug advertising and medication use: results of a national survey. *BMC Health Serv Res* 2011 Sep 23;11(1):232 [FREE Full text] [doi: [10.1186/1472-6963-11-232](https://doi.org/10.1186/1472-6963-11-232)] [Medline: [21942938](https://pubmed.ncbi.nlm.nih.gov/21942938/)]
5. Sinkinson M, Starc A. Ask your doctor? Direct-to-consumer advertising of pharmaceuticals. *The Review of Economic Studies* 2019;86(2):836-881 [FREE Full text] [doi: [10.1093/restud/rdy001](https://doi.org/10.1093/restud/rdy001)]
6. Health Tracking Poll, February 2019: Prescription Drugs. Kaiser Family Foundation. URL: <https://www.kff.org/health-costs/poll-finding/kff-health-tracking-poll-february-2019-prescription-drugs/> [accessed 2021-08-14]

7. Trust in healthcare. Edelman Trust Barometer. 2019. URL: https://www.edelman.com/sites/g/files/aatuss191/files/2019-04/2019_Edelman_Trust_Barometer_Global_Health_Report.pdf [accessed 2021-08-01]
8. About WeGO Health. WEGO Health. URL: <https://www.wegohealth.com/company-info/about/> [accessed 2021-06-12]
9. Enberg J. Influencer marketing in the age of COVID-19. eMarketer. 2020 Aug 24. URL: <https://www.emarketer.com/content/influencer-marketing-in-the-age-of-covid-19> [accessed 2021-05-10]
10. Zhao J, Ha S, Widdows R. The influence of social capital on knowledge creation in online health communities. *Inf Technol Manag* 2015 Jan 4;17(4):311-321 [FREE Full text] [doi: [10.1007/s10799-014-0211-3](https://doi.org/10.1007/s10799-014-0211-3)]
11. Lewinski AA, Fisher EB. Social interaction in type 2 diabetes computer-mediated environments: How inherent features of the channels influence peer-to-peer interaction. *Chronic Illn* 2016 Jun;12(2):116-144. [doi: [10.1177/1742395315601414](https://doi.org/10.1177/1742395315601414)] [Medline: [26304692](https://pubmed.ncbi.nlm.nih.gov/26304692/)]
12. Hawkins CM, DeLa OJ, Hung C. Social Media and the Patient Experience. *J Am Coll Radiol* 2016 Dec;13(12 Pt B):1615-1621. [doi: [10.1016/j.jacr.2016.09.006](https://doi.org/10.1016/j.jacr.2016.09.006)] [Medline: [27888950](https://pubmed.ncbi.nlm.nih.gov/27888950/)]
13. Attai DJ, Cowher MS, Al-Hamadani M, Schoger JM, Staley AC, Landercasper J. Twitter Social Media is an Effective Tool for Breast Cancer Patient Education and Support: Patient-Reported Outcomes by Survey. *J Med Internet Res* 2015 Jul 30;17(7):e188 [FREE Full text] [doi: [10.2196/jmir.4721](https://doi.org/10.2196/jmir.4721)] [Medline: [26228234](https://pubmed.ncbi.nlm.nih.gov/26228234/)]
14. Chandawarkar AA, Gould DJ, Grant Stevens W. The Top 100 Social Media Influencers in Plastic Surgery on Twitter: Who Should You Be Following? *Aesthet Surg J* 2018 Jul 13;38(8):913-917 [FREE Full text] [doi: [10.1093/asj/sjy024](https://doi.org/10.1093/asj/sjy024)] [Medline: [29518179](https://pubmed.ncbi.nlm.nih.gov/29518179/)]
15. Thomas K. Key opinion leaders supercharged by the internet: paid doctor and patient influencers on social media. *BMJ* 2019 May 31;365:l2336. [doi: [10.1136/bmj.l2336](https://doi.org/10.1136/bmj.l2336)] [Medline: [31152003](https://pubmed.ncbi.nlm.nih.gov/31152003/)]
16. Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: A concept analysis based on a literature review. *Patient Educ Couns* 2016 Dec;99(12):1923-1939. [doi: [10.1016/j.pec.2016.07.026](https://doi.org/10.1016/j.pec.2016.07.026)] [Medline: [27450481](https://pubmed.ncbi.nlm.nih.gov/27450481/)]
17. Amount of time US primary care physicians spent with each patient as of 2018. Statista. 2019. URL: <https://www.statista.com/statistics/250219/us-physicians-opinion-about-their-compensation/> [accessed 2021-07-18]
18. Naidu A. Factors affecting patient satisfaction and healthcare quality. *Int J Health Care Qual Assur* 2009;22(4):366-381. [doi: [10.1108/09526860910964834](https://doi.org/10.1108/09526860910964834)] [Medline: [19725209](https://pubmed.ncbi.nlm.nih.gov/19725209/)]
19. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties. *Ann Intern Med* 2016 Dec 06;165(11):753-760. [doi: [10.7326/M16-0961](https://doi.org/10.7326/M16-0961)] [Medline: [27595430](https://pubmed.ncbi.nlm.nih.gov/27595430/)]
20. Tajirian T, Stergiopoulos V, Strudwick G, Sequeira L, Sanches M, Kemp J, et al. The Influence of Electronic Health Record Use on Physician Burnout: Cross-Sectional Survey. *J Med Internet Res* 2020 Jul 15;22(7):e19274 [FREE Full text] [doi: [10.2196/19274](https://doi.org/10.2196/19274)] [Medline: [32673234](https://pubmed.ncbi.nlm.nih.gov/32673234/)]
21. Preece J. Sociability and usability in online communities: Determining and measuring success. *Behaviour & Information Technology* 2010 Nov 08;20(5):347-356 [FREE Full text] [doi: [10.1080/01449290110084683](https://doi.org/10.1080/01449290110084683)]
22. Willis E. The making of expert patients: the role of online health communities in arthritis self-management. *J Health Psychol* 2014 Dec;19(12):1613-1625 [FREE Full text] [doi: [10.1177/1359105313496446](https://doi.org/10.1177/1359105313496446)] [Medline: [23988679](https://pubmed.ncbi.nlm.nih.gov/23988679/)]
23. Willis E. Patients' self-efficacy within online health communities: facilitating chronic disease self-management behaviors through peer education. *Health Commun* 2016;31(3):299-307. [doi: [10.1080/10410236.2014.950019](https://doi.org/10.1080/10410236.2014.950019)] [Medline: [26325224](https://pubmed.ncbi.nlm.nih.gov/26325224/)]
24. Fox NJ, Ward KJ, O'Rourke AJ. The 'expert patient': empowerment or medical dominance? The case of weight loss, pharmaceutical drugs and the Internet. *Soc Sci Med* 2005 Mar;60(6):1299-1309. [doi: [10.1016/j.socscimed.2004.07.005](https://doi.org/10.1016/j.socscimed.2004.07.005)] [Medline: [15626525](https://pubmed.ncbi.nlm.nih.gov/15626525/)]
25. Frost JH, Massagli MP. Social uses of personal health information within PatientsLikeMe, an online patient community: what can happen when patients have access to one another's data. *J Med Internet Res* 2008 May 27;10(3):e15 [FREE Full text] [doi: [10.2196/jmir.1053](https://doi.org/10.2196/jmir.1053)] [Medline: [18504244](https://pubmed.ncbi.nlm.nih.gov/18504244/)]
26. Linn AJ, van Weert JCM, Gebeyehu BG, Sanders R, Diviani N, Smit EG, et al. Patients' Online Information-Seeking Behavior Throughout Treatment: The Impact on Medication Beliefs and Medication Adherence. *Health Commun* 2019 Nov;34(12):1461-1468. [doi: [10.1080/10410236.2018.1500430](https://doi.org/10.1080/10410236.2018.1500430)] [Medline: [30052088](https://pubmed.ncbi.nlm.nih.gov/30052088/)]
27. Wicks P, Keininger DL, Massagli MP, de la Loge C, Brownstein C, Isojärvi J, et al. Perceived benefits of sharing health data between people with epilepsy on an online platform. *Epilepsy Behav* 2012 Jan;23(1):16-23 [FREE Full text] [doi: [10.1016/j.yebeh.2011.09.026](https://doi.org/10.1016/j.yebeh.2011.09.026)] [Medline: [22099528](https://pubmed.ncbi.nlm.nih.gov/22099528/)]
28. Willis E, Royne MB. Online Health Communities and Chronic Disease Self-Management. *Health Commun* 2017 Mar 24;32(3):269-278. [doi: [10.1080/10410236.2016.1138278](https://doi.org/10.1080/10410236.2016.1138278)] [Medline: [27218836](https://pubmed.ncbi.nlm.nih.gov/27218836/)]
29. Willis E. Applying the Health Belief Model to Medication Adherence: The Role of Online Health Communities and Peer Reviews. *J Health Commun* 2018;23(8):743-750. [doi: [10.1080/10810730.2018.1523260](https://doi.org/10.1080/10810730.2018.1523260)] [Medline: [30280998](https://pubmed.ncbi.nlm.nih.gov/30280998/)]
30. De Choudhury M, Morris MR, White RW. Seeking and sharing health information online: Comparing search engines and social media. 2014 Presented at: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2014; Toronto, Canada p. 1365-1376 URL: <https://doi.org/10.1145/2556288.2557214> [doi: [10.1145/2556288.2557214](https://doi.org/10.1145/2556288.2557214)]

31. 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](https://doi.org/10.2196/jmir.1933)] [Medline: [23615206](https://pubmed.ncbi.nlm.nih.gov/23615206/)]
32. Millard T, Agius PA, McDonald K, Slavin S, Girdler S, Elliott JH. The Positive Outlook Study: A Randomised Controlled Trial Evaluating Online Self-Management for HIV Positive Gay Men. *AIDS Behav* 2016 Sep;20(9):1907-1918. [doi: [10.1007/s10461-016-1301-5](https://doi.org/10.1007/s10461-016-1301-5)] [Medline: [26896121](https://pubmed.ncbi.nlm.nih.gov/26896121/)]
33. Trudeau KJ, Pujol LA, DasMahapatra P, Wall R, Black RA, Zacharoff K. A randomized controlled trial of an online self-management program for adults with arthritis pain. *J Behav Med* 2015 Jun;38(3):483-496 [FREE Full text] [doi: [10.1007/s10865-015-9622-9](https://doi.org/10.1007/s10865-015-9622-9)] [Medline: [25686538](https://pubmed.ncbi.nlm.nih.gov/25686538/)]
34. Kanyan A, Andrew JV, Ali JK, Beti MM. Building Customer Relationship for Gaining Customer Loyalty in the Pharmaceutical Industry. *JOAMS* 2015:319-322. [doi: [10.12720/joams.3.4.319-322](https://doi.org/10.12720/joams.3.4.319-322)]
35. Wolin B. Big data, pharmaceutical marketing and healthcare. *Digital & Social Media Marketing* 2014;2(1):35-39.
36. Enyinda CI, Ogbuehi AO, Mbah CH. Building pharmaceutical relationship marketing and social media impact. *IJPHM* 2018 Jun 29;12(2):198-230 [FREE Full text] [doi: [10.1108/ijphm-02-2017-0007](https://doi.org/10.1108/ijphm-02-2017-0007)]
37. Dhanesh GS, Duthler G. Relationship management through social media influencers: Effects of followers' awareness of paid endorsement. *Public Relations Review* 2019 Sep;45(3):101765 [FREE Full text] [doi: [10.1016/j.pubrev.2019.03.002](https://doi.org/10.1016/j.pubrev.2019.03.002)]
38. Delbaere M, Michael B, Phillips BJ. Social media influencers: A route to brand engagement for their followers. *Psychology & Marketing* 2020 Oct 11;38(1):101-112 [FREE Full text] [doi: [10.1002/mar.21419](https://doi.org/10.1002/mar.21419)]
39. Enke N, Borchers NS. Social Media Influencers in Strategic Communication: A Conceptual Framework for Strategic Social Media Influencer Communication. *International Journal of Strategic Communication* 2019 Sep 04;13(4):261-277 [FREE Full text] [doi: [10.1080/1553118x.2019.1620234](https://doi.org/10.1080/1553118x.2019.1620234)]
40. Boerman SC, Willemsen LM, Van Der Aa EP. "This Post Is Sponsored": Effects of sponsorship disclosure on persuasion knowledge and electronic word of mouth in the context of Facebook. *Journal of Interactive Marketing* 2017 May;38:82-92 [FREE Full text] [doi: [10.1016/j.intmar.2016.12.002](https://doi.org/10.1016/j.intmar.2016.12.002)]
41. Drosch B. What retailers need to know about influencer marketing. *eMarketer*. 2019. URL: <https://www.emarketer.com/content/what-retailers-need-to-know-about-influencer-marketing> [accessed 2021-07-15]
42. Lou C, Yuan S. Influencer Marketing: How Message Value and Credibility Affect Consumer Trust of Branded Content on Social Media. *Journal of Interactive Advertising* 2019 Feb 12;19(1):58-73 [FREE Full text] [doi: [10.1080/15252019.2018.1533501](https://doi.org/10.1080/15252019.2018.1533501)]
43. Casaló L, Flavián C, Ibáñez-Sánchez S. Influencers on Instagram: Antecedents and consequences of opinion leadership. *Journal of Business Research* 2020 Sep;117:510-519 [FREE Full text] [doi: [10.1016/j.jbusres.2018.07.005](https://doi.org/10.1016/j.jbusres.2018.07.005)]
44. Schouten AP, Janssen L, Verspaget M. Celebrity vs. Influencer endorsements in advertising: the role of identification, credibility, and Product-Endorser fit. *International Journal of Advertising* 2019 Jul 02;39(2):258-281 [FREE Full text] [doi: [10.1080/02650487.2019.1634898](https://doi.org/10.1080/02650487.2019.1634898)]
45. Kolo C, Haumer F. Social media celebrities as influencers in brand communication: An empirical study on influencer content, its advertising relevance and audience expectations. *Digital & Social Media Marketing* 2018;6(3):A.
46. Costa Do Nascimento T, Dias Campos R, Suarez M. Experimenting, partnering and bonding: a framework for the digital influencer-brand endorsement relationship. *Journal of Marketing Management* 2020 Jul 20;36(11-12):1009-1030 [FREE Full text] [doi: [10.1080/0267257x.2020.1791933](https://doi.org/10.1080/0267257x.2020.1791933)]
47. Jiménez-Castillo D, Sánchez-Fernández R. The role of digital influencers in brand recommendation: Examining their impact on engagement, expected value and purchase intention. *International Journal of Information Management* 2019 Dec;49:366-376 [FREE Full text] [doi: [10.1016/j.ijinfomgt.2019.07.009](https://doi.org/10.1016/j.ijinfomgt.2019.07.009)]
48. Booth N, Matic JA. Mapping and leveraging influencers in social media to shape corporate brand perceptions. *Corporate Communications: An International Journal* 2011 Aug 09;16(3):184-191 [FREE Full text] [doi: [10.1108/13563281111156853](https://doi.org/10.1108/13563281111156853)]
49. For industry: Using social media. Food and Drug Administration. 2014. URL: <https://www.fda.gov/aboutfda/centersoffices/officeofmedicalproductsandtobacco/cder/ucm397791.htm> [accessed 2021-09-19]
50. Presenting quantitative efficacy and risk information in direct-to-consumer promotional labeling and advertisements. Food and Drug Administration. URL: <https://www.fda.gov/media/117573/download> [accessed 2021-08-19]
51. Disclosures 101 for social media influencers. Federal Trade Commission. 2019. URL: <https://www.ftc.gov/tips-advice/business-center/guidance/disclosures-101-social-media-influencers> [accessed 2021-10-05]
52. Pharmaceuticals maneuver advertising restrictions with influencers. *MediaKix*. 2021. URL: <https://mediakix.com/blog/pharmaceutical-influencer-marketing/> [accessed 2021-12-01]
53. Warning letters and notice of violation letters to pharmaceutical companies. Food and Drug Administration. 2015. URL: <http://wayback.archive-it.org/7993/20170112061359/http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/EnforcementActivitiesbyFDA/WarningLettersandNoticeofViolationLetterstoPharmaceuticalCompanies/UCM457961.pdf> [accessed 2021-06-05]
54. Warning letters and notice of violation letters to pharmaceutical companies. Food and Drug Administration. 2015. URL: <http://wayback.archive-it.org/7993/20170112061400/http://www.fda.gov/downloads/Drugs/>

- [GuidanceComplianceRegulatoryInformation/EnforcementActivitiesbyFDA/WarningLettersandNoticeofViolationLetterstoPharmaceuticalCompanies/UCM457965.pdf](#) [accessed 2021-09-18]
55. What is a micro-influencer? MediaKix. 2019. URL: <https://mediakix.com/blog/what-are-micro-influencers/> [accessed 2021-10-01]
 56. Zuppello S. The latest Instagram influencer frontier? Medical promotions. Vox. 2019. URL: <https://www.vox.com/the-goods/2019/2/15/18211007/medical-sponcon-instagram-influencer-pharmaceutical> [accessed 2021-07-29]
 57. Schwartz LM, Woloshin S. Medical Marketing in the United States, 1997-2016. JAMA 2019 Jan 01;321(1):80-96 [FREE Full text] [doi: [10.1001/jama.2018.19320](https://doi.org/10.1001/jama.2018.19320)] [Medline: [30620375](https://pubmed.ncbi.nlm.nih.gov/30620375/)]
 58. Moynihan R. Key opinion leaders: independent experts or drug representatives in disguise? BMJ 2008 Jun 21;336(7658):1402-1403 [FREE Full text] [doi: [10.1136/bmj.39575.675787.651](https://doi.org/10.1136/bmj.39575.675787.651)] [Medline: [18566074](https://pubmed.ncbi.nlm.nih.gov/18566074/)]
 59. Kennedy K. How many of your key opinion leaders are also digital opinion leaders? Creation.co. 2016. URL: <https://creation.co/knowledge/how-many-of-your-key-opinion-leaders-are-also-digital-opinion-leaders/> [accessed 2021-08-01]
 60. Physician influencers. Klear. 2021. URL: <https://klear.com/influencers/Physicians> [accessed 2021-09-18]
 61. Anusic I. Pharma Marketing Cheat Sheet: 7 Social Media Hacks. Mediatoolkit. 2021. URL: <https://www.mediatoolkit.com/blog/7-social-media-hacks-a-cheat-sheet-for-pharma-marketers/> [accessed 2021-11-04]
 62. Johnson E. Digital influencer marketing for pharma brands. 2019. URL: <https://www.convergencepointmedia.com/digital-influencer-marketing-for-pharma-brands/> [accessed 2021-10-01]
 63. Robinson R. Patient influencers are transforming pharma marketing strategies. PharmaVOICE. 2019. URL: <https://www.pharmavoice.com/article/2019-11-patient-influencers/> [accessed 2021-10-02]
 64. Adams C. Direct-to-Consumer Advertising of Prescription Drugs Can Inform the Public and Improve Health. JAMA Oncol 2016 Nov 01;2(11):1395-1396 [FREE Full text] [doi: [10.1001/jamaoncol.2016.2443](https://doi.org/10.1001/jamaoncol.2016.2443)] [Medline: [27532690](https://pubmed.ncbi.nlm.nih.gov/27532690/)]
 65. Tyrawski J, DeAndrea DC. Pharmaceutical companies and their drugs on social media: a content analysis of drug information on popular social media sites. J Med Internet Res 2015 Jun 01;17(6):e130 [FREE Full text] [doi: [10.2196/jmir.4357](https://doi.org/10.2196/jmir.4357)] [Medline: [26032738](https://pubmed.ncbi.nlm.nih.gov/26032738/)]
 66. Calfee JE. Public Policy Issues in Direct-to-Consumer Advertising of Prescription Drugs. Journal of Public Policy & Marketing 2018 Oct 02;21(2):174-193. [doi: [10.1509/jppm.21.2.174.17580](https://doi.org/10.1509/jppm.21.2.174.17580)]
 67. Myers SD, Royne MB, Deitz GD. Direct-to-Consumer Advertising: Exposure, Behavior, and Policy Implications. Journal of Public Policy & Marketing 2011 Apr 01;30(1):110-118 [FREE Full text] [doi: [10.1509/jppm.30.1.110](https://doi.org/10.1509/jppm.30.1.110)]
 68. Friestad M, Wright P. The Persuasion Knowledge Model: How People Cope with Persuasion Attempts. Journal of Consumer Research 1994 Jun;21(1):1-31 [FREE Full text] [doi: [10.1086/209380](https://doi.org/10.1086/209380)]
 69. Grigsby JL. Fake Ads: The Influence of Counterfeit Native Ads on Brands and Consumers. Journal of Promotion Management 2020 Jan 29;26(4):569-592 [FREE Full text] [doi: [10.1080/10496491.2020.1719958](https://doi.org/10.1080/10496491.2020.1719958)]
 70. Gurkaynak G, Kama CU. Navigating the uncharted risks of covert advertising in influencer marketing. Business Law Review 2018;39(1):17-19.
 71. Campbell C, Grimm PE. The Challenges Native Advertising Poses: Exploring Potential Federal Trade Commission Responses and Identifying Research Needs. Journal of Public Policy & Marketing 2018 Dec 19;38(1):110-123 [FREE Full text] [doi: [10.1177/0743915618818576](https://doi.org/10.1177/0743915618818576)]
 72. Taiminen K, Luoma-aho V, Tolvanen K. The transparent communicative organization and new hybrid forms of content. Public Relations Review 2015 Dec;41(5):734-743 [FREE Full text] [doi: [10.1016/j.pubrev.2015.06.016](https://doi.org/10.1016/j.pubrev.2015.06.016)]
 73. Schauster EE, Ferrucci P, Neill MS. Native Advertising Is the New Journalism. American Behavioral Scientist 2016 Jul 28;60(12):1408-1424 [FREE Full text] [doi: [10.1177/0002764216660135](https://doi.org/10.1177/0002764216660135)]
 74. Zarzosa J, Fischbach S. Native Advertising: Trickery or Technique? An Ethics Project and Debate. Marketing Education Review 2017 Apr 03;27(2):104-108 [FREE Full text] [doi: [10.1080/10528008.2017.1304161](https://doi.org/10.1080/10528008.2017.1304161)]
 75. Benetoli A, Chen TF, Aslani P. How patients' use of social media impacts their interactions with healthcare professionals. Patient Educ Couns 2018 Mar;101(3):439-444. [doi: [10.1016/j.pec.2017.08.015](https://doi.org/10.1016/j.pec.2017.08.015)] [Medline: [28882545](https://pubmed.ncbi.nlm.nih.gov/28882545/)]
 76. Vassilev I, Rogers A, Kennedy A, Koetsenruijter J. The influence of social networks on self-management support: a metasynthesis. BMC Public Health 2014 Jul 15;14:719 [FREE Full text] [doi: [10.1186/1471-2458-14-719](https://doi.org/10.1186/1471-2458-14-719)] [Medline: [25023948](https://pubmed.ncbi.nlm.nih.gov/25023948/)]
 77. Raine L, Perrin A. Key findings about Americans' declining trust in government and each other. Pew Research Center. 2019. URL: <https://www.pewresearch.org/fact-tank/2019/07/22/key-findings-about-americans-declining-trust-in-government-and-each-other/> [accessed 2022-02-16]
 78. Edelman Trust Barometer 2021. Edelman. 2021. URL: <https://www.edelman.com/sites/g/files/aatuss191/files/2021-01/2021-edelman-trust-barometer.pdf> [accessed 2021-10-05]
 79. Huang EC, Pu C, Chou Y, Huang N. Public Trust in Physicians-Health Care Commodification as a Possible Deteriorating Factor: Cross-sectional Analysis of 23 Countries. Inquiry 2018;55:46958018759174 [FREE Full text] [doi: [10.1177/0046958018759174](https://doi.org/10.1177/0046958018759174)] [Medline: [29502479](https://pubmed.ncbi.nlm.nih.gov/29502479/)]
 80. The ROI of influencer marketing compared to other marketing activities. eMarketer. 2020. URL: <https://www.emarketer.com/chart/243671/>

- [roi-of-influencer-marketing-compared-other-marketing-activities-according-us-marketers-june-2020-of-respondents](#) [accessed 2021-11-04]
81. Feick LL, Price LL. The Market Maven: A Diffuser of Marketplace Information. *Journal of Marketing* 2018 Dec 13;51(1):83-97 [FREE Full text] [doi: [10.1177/002224298705100107](#)]
 82. O'Sullivan SR. The Market Maven Crowd: Collaborative Risk-Aversion and Enhanced Consumption Context Control in an Illicit Market. *Psychology & Marketing* 2015 Jan 27;32(3):285-302 [FREE Full text] [doi: [10.1002/mar.20780](#)]
 83. Grindrod K, Forgione A, Tsuyuki RT, Gavura S, Giustini D. Pharmacy 2.0: a scoping review of social media use in pharmacy. *Res Social Adm Pharm* 2014;10(1):256-270 [FREE Full text] [doi: [10.1016/j.sapharm.2013.05.004](#)] [Medline: [23810653](#)]
 84. Alves TL, Mantel-Teeuwisse AK, Paschke A, Leufkens HGM, Puil L, Poplavska E, et al. Unbranded advertising of prescription medicines to the public by pharmaceutical companies. *Cochrane Systematic Review* 2017(7):1 [FREE Full text] [doi: [10.1002/14651858.CD012699](#)]
 85. Huh J, Suzuki-Lambrecht Y, Lueck J, Gross M. Presentation Matters: Comparison of Cognitive Effects of DTC Prescription Drug Advergates, Websites, and Print Ads. *Journal of Advertising* 2015 Feb 04;44(4):360-374 [FREE Full text] [doi: [10.1080/00913367.2014.1003666](#)]
 86. Byrne E, Kearney J, MacEvilly C. The Role of Influencer Marketing and Social Influencers in Public Health. *Proc Nutr Soc* 2017 Sep 29;76(OCE3):E103 [FREE Full text] [doi: [10.1017/s0029665117001768](#)]
 87. Wei M, Delbaere M. Do consumers perceive their doctors as influenced by pharmaceutical marketing communications? *International Journal of Pharmaceutical and Healthcare Marketing* 2015;9(4):330-348. [doi: [10.1108/ijphm-06-2014-0033](#)]
 88. Slobodzian A, Delbaere M. Covert Persuasion Attempts: Do People Notice? In: Krey N, Rossi P, editors. *Back to the Future: Using Marketing Basics to Provide Customer Value*. Cambridge: Springer; 2017.

Abbreviations

FDA: Food and Drug Administration

FTC: Federal Trade Commission

Edited by A Mavragani; submitted 06.04.21; peer-reviewed by MA Bahrami, E Chiauzzi; comments to author 02.07.21; revised version received 22.07.21; accepted 16.12.21; published 01.03.22.

Please cite as:

Willis E, Delbaere M

Patient Influencers: The Next Frontier in Direct-to-Consumer Pharmaceutical Marketing

J Med Internet Res 2022;24(3):e29422

URL: <https://www.jmir.org/2022/3/e29422>

doi: [10.2196/29422](#)

PMID: [35230241](#)

©Erin Willis, Marjorie Delbaere. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 01.03.2022. 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.

Viewpoint

How Can Research on Artificial Empathy Be Enhanced by Applying Deepfakes?

Hsuan-Chia Yang^{1,2,3,4}, PhD; Annisa Ristya Rahmanti^{1,2,5}, MS; Chih-Wei Huang^{2*}, PhD; Yu-Chuan Jack Li^{1,2,3,6*}, MD, PhD, FACMI

¹Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

²International Center for Health Information Technology, Taipei Medical University, Taipei, Taiwan

³Research Center of Big Data and Meta-analysis, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

⁴Clinical Big Data Research Center, Taipei Medical University Hospital, Taipei, Taiwan

⁵Department of Health Policy Management, Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

⁶Department of Dermatology, Wanfang Hospital, Taipei, Taiwan

* these authors contributed equally

Corresponding Author:

Yu-Chuan Jack Li, MD, PhD, FACMI

Graduate Institute of Biomedical Informatics

College of Medical Science and Technology

Taipei Medical University

No 172-1, Sec 2 Keelung Rd

Taipei, 106

Taiwan

Phone: 886 966 546 813

Fax: 886 2 6638 0233

Email: jack@tmu.edu.tw

Abstract

We propose the idea of using an open data set of doctor-patient interactions to develop artificial empathy based on facial emotion recognition. Facial emotion recognition allows a doctor to analyze patients' emotions, so that they can reach out to their patients through empathic care. However, face recognition data sets are often difficult to acquire; many researchers struggle with small samples of face recognition data sets. Further, sharing medical images or videos has not been possible, as this approach may violate patient privacy. The use of deepfake technology is a promising approach to deidentifying video recordings of patients' clinical encounters. Such technology can revolutionize the implementation of facial emotion recognition by replacing a patient's face in an image or video with an unrecognizable face—one with a facial expression that is similar to that of the original. This technology will further enhance the potential use of artificial empathy in helping doctors provide empathic care to achieve good doctor-patient therapeutic relationships, and this may result in better patient satisfaction and adherence to treatment.

(*J Med Internet Res* 2022;24(3):e29506) doi:[10.2196/29506](https://doi.org/10.2196/29506)

KEYWORDS

artificial empathy; deepfakes; doctor-patient relationship; face emotion recognition; artificial intelligence; facial recognition; facial emotion recognition; medical images; patient; physician; therapy

Introduction

Good doctor-patient communication is one of the key requirements of building a successful, therapeutic doctor-patient relationship [1]. This type of communication enables physicians to provide better-quality care that may impact patients' health. Studies on good doctor-patient communication have demonstrated a strong positive correlation between physician communication skills and patient satisfaction, which is likely

associated with patients' adherence to treatment; their experience of care; and, consequently, improved clinical outcomes [2-5].

We acknowledge the importance of good doctor-patient communication; doctors must understand patients' perspectives through verbal conversation and nonverbal behaviors (eg, posture, gesture, eye contact, facial expression, etc) [6,7]. Establishing communication involving nonverbal messages is very important in building a good doctor-patient relationship because such communication conveys more expressive and

meaningful messages than those conveyed in a verbal conversation [8]. One research study indicates that nonverbal messages contribute to up to 90% of messages delivered in human interactions [6]. Another study also estimates that more than half of outpatient clinic patients believe that establishing positive nonverbal behaviors indicates that a doctor is more attentive toward their patient and thus results in better patient satisfaction and adherence to treatment [8].

Although several studies have reported that human nonverbal behaviors are significantly associated with patient satisfaction and compliance to a treatment plan, physicians are often clueless about nonverbal messages [6]. Doctors should be more aware of their nonverbal behaviors because patients are cognizant of them. Doctors also need to recognize and evaluate patients' nonverbal behaviors and their own nonverbal behaviors toward patients.

Artificial intelligence (AI) offers great potential for exploring nonverbal communication in doctor-patient encounters [9]. For example, AI may help a doctor become more empathic by analyzing human facial expressions through emotion recognition. Once an emotionally intelligent AI identifies an emotion, it can guide a doctor to express more empathy based on each patient's unique emotional needs [10].

Empathy refers to the ability to understand or feel what another person is experiencing, and showing empathy may lead to better behavioral outcomes [9]. Empathy can be learned, and the use of AI technology introduces a promising approach to incorporating artificial empathy in the doctor-patient therapeutic relationship [11]. However, human emotions are very complex. An emotionally intelligent AI should learn a range of emotions (ie, those that patients experience) from facial expressions, voices, and physiological signals to empathize with human emotions [12]. These emotions can be captured by using various modalities, such as video, audio, text, and physiological signals [13].

Among all forms of human communication channels, facial expressions are recognized as the most essential and influential [14-16]. The human face can express various thoughts, emotions, and behaviors [15]. It can convey important aspects in human interpersonal communication and nonverbal expressions in social interactions [17,18]. Compared to the amount of information that can be conveyed via emotion recognition technology, facial expressions convey 55% of the emotional expression transmitted in multimodal human interactions, whereas verbal information, text communication, and communication via physiological signals only convey 20%, 15%, and 10% of the total information in interactions, respectively [19].

Many researchers have been studying facial expressions by using automatic facial emotion recognition (FER) to gain a better understanding of the human emotions linked with empathy [20-24]. They have proposed various machine learning algorithms, such as support vector machines, Bayesian belief networks, and neural network models, for recognizing and describing emotions based on observed facial expressions recorded on images or videos [20-22]. Although mounting literature has been introduced on machine learning and deep

learning for automatically extracting emotions from the human face, developing a highly accurate FER system requires a lot of training data and a high-quality computational system [21]. In addition, the data set must include diverse facial views in terms of angles, frame rates, races, and genders, among others [21].

Many public data sets are available for FER [25]. However, most public data sets are not sufficient for supporting doctor-patient interactions. Creating our own medical data sets is also not possible, since this process is expensive and time consuming [26]. Moreover, researchers often struggle with acquiring sufficient data for training a face recognition model due to privacy concerns. Data sharing and the pooling of medical images or videos are not even possible, as these approaches may violate patient privacy. Herein, we present our study on the emerging AI is known as *deepfakes*—a technology that enables face deidentification for recorded videos of patients' clinical encounters. This technology can revolutionize FER by replacing patients' faces in images or videos with an unrecognizable face, thereby anonymizing patients. This could protect patients' privacy when it comes to clinical encounter videos and allow medical video data sharing to become more feasible. Moreover, using an open clinical encounter video data set can also promote more in-depth research within the academic community. Thus, deepfake technology will further enhance the clinical application of artificial empathy for medical application purposes.

Methods

Human FER

Human FER plays a significant role in understanding people's nonverbal ways of communicating with others [19]. It has attracted the interest of scientific populations in various fields due to its superiority among other forms of emotion recognition [22]. As it is not only limited to human-computer interactions or human-robot interactions, facial expression analysis has become a popular research topic in various health care areas, such as the diagnosis or assessment of cognitive impairment (eg, autism spectrum disorders in children), depression monitoring, pain monitoring in Parkinson Disease, and clinical communication in doctor-patient consultations [27].

The main objective of FER is to accurately classify various facial expressions according to a person's emotional state [21]. The classical FER approach is usually divided into the following three major stages: (1) facial feature detection, (2) feature extraction, and (3) emotion recognition [21,28]. However, traditional FER has been reported to be unable to extract facial expressions in an uncontrolled environment with diverse facial views [21,28]. On the other hand, a recent study using a deep learning-based FER approach has successfully achieved superior accuracy over that of traditional FER [20-22].

Deepfake Technology

The rapid growth of computer vision and deep learning technology has driven the recently emerged phenomena of deepfakes (*deep learning* and *fake*), which can automatically forge images and videos that humans cannot easily recognize

[29-31]. In addition, deepfake techniques offer the possibility of generating unrecognizable images of a person’s face and altering or swapping a person’s face in existing images and videos with another face that exhibits the same expressions as the original face [29]. Various deepfake attempts have been used for negative purposes, such as creating controversial content related to celebrities, politicians, companies, and even individuals to damage their reputation [30]. Although the harmful effects of deepfake technology have raised public concerns, there are also advantages to using this technology. For example, it can provide privacy protection in some critical medical applications, such as face deidentification for patients [32]. Further, although deepfake technology can easily manipulate the low-level semantics of visual and audio features, a recent study suggested that it might be difficult for deepfake technology to forge the high-level semantic features of human emotions [31].

Deepfake technology is mainly developed by using deep learning—an AI-based method that can be used to train deep networks [29]. The popular approach to implementing deepfake techniques is based on the generative adversarial network (GAN) model [33,34]. There are several types and examples of deepfakes, such as photo deepfakes, audio deepfakes, video deepfakes, and audio-video deepfakes.

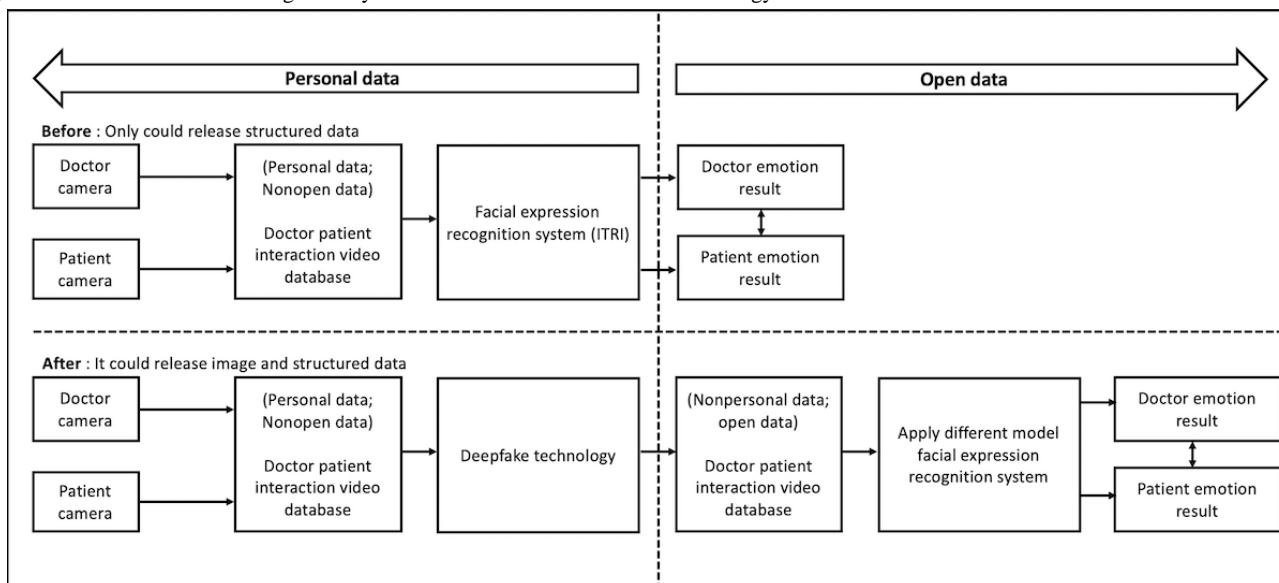
Data Set

To simulate how deepfake technology enables face deidentification for recorded videos of doctor-patient clinical encounters, we recruited 348 adult patients and 4 doctors from Taipei Municipal Wanfang Hospital and Taipei Medical University Hospital from March to December 2019. After excluding video data from 21 patients due to video damage, we collected video data from 327 patients. The data set focused on the interactions between doctors and patients in dermatology outpatient clinics. The subjects in the data set are all from the Taiwanese population.

The FER System in the Deepfake Model Setup

Figure 1 illustrates the workflow of the FER system before and after proposing deepfake technology. First, we created synchronized recordings by using 2 cameras to capture doctor-patient interactions in the dermatology outpatient clinic. We assumed that the face was the most relevant and accessible channel for nonverbal communication in health care [6]. Therefore, we then used a facial expression recognition system developed by the Industrial Technology Research Institute to detect emotions and analyze the emotional changes of the doctors and patients across time. This facial expression recognition system has been deployed using training data from 28,710 Asian face images and has an accuracy of 95% for the extended Cohn-Kanade data set [35].

Figure 1. The facial emotion recognition system workflow. ITRI: Industrial Technology Research Institute.



We identified facial expressions by using the main points of an individual’s face (eg, eyes, eyebrows, the tip of the nose, lip corners, etc) to track facial movement. This allowed us to observe the emotional experiences of the doctors and patients when they expressed the following seven facial expressions: anger, disgust, fear, neutral, happiness, sadness, and surprise. The system then provided a summary of the emotional changes of both the doctors and the patients with a temporal resolution of up to 1 second. Additionally, our model managed to filter out any irrelevant face targets (ie, faces other than those of the doctors and patients). Finally, the summary results of the doctor and patient emotion analyses were used as a reference data set

to develop artificial empathy. The system then created recommendations, so that doctors could provide an immediate response based on patients’ situations.

It should be noted however that our artificial empathy reference data training set was built by using limited face recognition data sets. Therefore, we tried to improve the model by proposing the use of open data from a clinical encounter video manipulated by deepfake technology, which can enable medical data sharing without violating patient privacy. Furthermore, these open data allowed us to connect with real-world clinical encounter video data sets, so that we could use different model facial expression

recognition systems to analyze patients' and doctors' emotional experiences (Figure 1).

Ethics Approval

Our study was approved by Taipei Medical University (TMU)-Joint Institutional Review Board (TMU-JIRB No: N201810020).

Results

The clinical encounter video—the source of our face recognition data set—consists of video data from 327 patients—208 female patients and 119 male patients (age: mean 51, SD 19.06 years). The average consultation time on the recorded video was 4.61 (SD 3.04) minutes; the longest duration of a consultation was 25.55 minutes, and the shortest was 0.33 minutes. Our artificial empathy algorithm was developed by using FER algorithms. This algorithm learned a range of patient emotions by analyzing expressions, so that doctors could provide an immediate response based on patients' emotional experiences. In general, this FER system achieved a mean detection rate of >80% on real-world data.

Our face recognition data set for artificial empathy was solely based on basic emotions. The system evaluation reported expressions of anger, happiness, disgust, and sadness, which were more likely to be expressed by the doctors than by the patients ($P<.001$). Moreover, patients also tended to more commonly express neutral emotions and surprise when compared to doctors ($P<.001$). The overall emotions of the doctors were dominated by emotions of sadness (expressions: 8580/17,397, 49.3%), happiness (expressions: 7541/17,397, 43.3%), anger (expressions: 629/17,397, 3.6%), surprise (expressions: 436/17,397, 2.5%), and disgust (expressions: 201/17,397, 1.2%), whereas the emotions of patients consisted of happiness (expressions: 5766/12,606, 45.7%), sadness (expressions: 5773/12,606, 45.8%), surprise (expressions: 890/12,606, 7.1%), and anger (expressions: 126/12,606, 0.9%). Figure 2 illustrates the emotional expressions of both doctors and patients. The system used the results of the emotion analysis to remind the doctors to change their behaviors according to patients' situations, so that the patients felt like the doctors understood their emotions and situations.

The original face recognition data set consists of personal data (ie, patients' faces). However, we can only release the results of the emotional expression analysis as a reference for the development of artificial empathy. As noted previously, our approach only involved using a small amount of training data (only Asian face images). Therefore, to improve model performance, we need to anonymize the clinical interaction video by performing face deidentification. Face deidentification allows us to share our face recognition data set as open data for clinical research. To enable face image data sharing, a researcher can perform traditional face deidentification techniques, such as masking an image by covering a patient's face region with a colored box (Figure 3).

Of note however, as our research aims to develop artificial empathy to support good doctor-patient relationships, the masking method cannot be performed, as it is very difficult to validate masked images with the results of an emotion expression analysis. Deepfake technology offers a method for swapping a patient's original face with another from an open-source face data set to generate an unrecognizable image with similar expressions and attributes to those of the original face image. This face swapping method can be adopted for use with the face recognition reference data set for our artificial empathy algorithm to avoid violating patient privacy and ethical concerns. We adopted video deepfake technology based on face swapping (Figure 3), which was proposed in the first order motion model for image animation [36]. This approach involved adopting a novel deep learning framework for image animation known as *Monkey-Net* and modifying it by using a set of self-learned key points combined with local affine transformations [36]. This framework enables a dense motion transfer network to generate a video in which the source image is animated according to a given driving video sequence with complex motions [36]. Unlike the original GAN model, which relied on costly ground-truth pretrained models that resulted in the poor generation quality of image or video outputs, the first order motion model for image animation can handle high-resolution data sets with profile images and can thus become a reference benchmark model for our face recognition data set.

Figure 2. Screenshots of the recorded video simulation of the doctor-patient relationship in the dermatology outpatient clinic.



Figure 3. Comparison between traditional face deidentification and face swapping by using deepfake technology on an image of a patient's face.



Discussion

Principal Findings

Our FER study revealed how doctors more commonly express emotions like anger, happiness, disgust, and sadness than patients. Because nonverbal messages like facial expressions contribute the most to the messages delivered in human interactions, doctors need to be more careful when expressing their emotions during clinical interactions. For example, doctors should never be used to expressing anger, disgust, or other negative emotions that represent poor communication skills, as this may ruin treatment goals and result in frustration for both patients and health care practitioners [6].

Positive emotions (eg, happiness) represent good communication skills, as they may help people understand how another person feels and what they think and allow people to understand each other better [37]. Furthermore, positive emotions can help build patients' trust in their doctors [38]. Trust from a patient's perspective refers to the acceptance of a vulnerable situation in which patients believe that doctors will provide adequate and fair medical care to help them based on their needs [39]. When patients trust their doctors, they are more likely to share valid and reliable information related to their condition, acknowledge health problems more readily, comprehend medical information efficiently, and comply with treatment plans accordingly [39]. They also tend to seek preventive care earlier and return for follow-up care, which may prevent further disease complications [39].

In addition to physicians' medical knowledge and clinical skills, patients' perceptions of physicians' ability to provide adequate information, actively listen, and empathize are believed to be associated with patient satisfaction and trust [3]. A physician's capability to exhibit effective communication skills and provide empathic care is beneficial for patients in terms of improving good doctor-patient relationships and for the physicians themselves, as these factors can increase job performance satisfaction and lower the risk of stress and physical burnout among physicians [40]. Empathic care may also reduce the rate of medical errors and help to avoid conflict with patients [38].

We believe that our FER system and face recognition data set can serve as a decision support system that can guide doctors when a patient requires special attention for achieving therapeutic goals. For example, if doctors express a negative facial expression (eg, anger, disgust, and sadness), the system will remind them to change their facial expressions. Moreover, if a patient also expresses a negative facial expression, the system will suggest that the doctor should use a different approach to accommodate the patient's emotional situation. Based on our results, the major shortcoming that we need to address is that FER technology relies on the quality of data training and the quantity of training data [26,32]. We believe that in the future, we can improve the system's precision and accuracy by collecting more data from more subjects with various sociodemographic backgrounds. This is only possible if we adopt deepfake technology (eg, GANs), which can learn the facial features of a human face on images and videos and replace it with another person's face [41]. Thus, deepfake technology can replace a patient's face image and create fake face images with similar facial expressions in videos. With the use of deepfake technology, the recorded video database of outpatient doctor-patient interactions will become more accessible. Applying deepfakes to deidentify FER data sets may benefit the development of artificial empathy, as this approach may not violate the privacy and security of interpersonal situations.

Similar to our study, a recent study reported using deepfake technology to generate open-source, high-quality medical video data sets of Parkinson disease examination videos to deidentify subjects [32]. This study also applied the face swapping technique and real-time multi-person system to detect human motion key points based on open-source videos from the Deep Fake Detection data set [32]. Meanwhile, our approach involved using a self-supervised formulation consisting of self-learned key points combined with local affine transformations [36]. We believe that this self-learned model could preserve the represented emotional states of people in the original face recognition data set.

Our study has some limitations. First, our approach only involved using a single information modality—video deepfakes—which could have resulted in inaccurate emotion classification. In the future, we can combine both video and audio deepfakes to better represent the emotional states of a target person. Second, moral and ethical concerns need to be considered when using deepfake technology for the deidentification of medical data sets. However, our study highlighted the positive ways of using deepfakes for privacy protection when using face recognition data sets in medical settings. Thus, instead of raising an ethical problem, this study will help prevent the use of deepfakes for malicious purposes and encourage their use in medical applications.

Conclusion

We propose using an open data set of clinical encounter videos as a reference data training set to develop artificial empathy

based on an FER system, given that FER technologies rely on extensive data training. Yet, due to privacy concerns, it has always been difficult for researchers to acquire a face recognition data set. Therefore, we suggest the adoption of deepfakes. Deepfake technology can deidentify faces in images or videos and manipulate them so that the proper target face becomes unrecognizable, thereby preventing the violation of patient privacy. Such technology can also generate the same facial expressions as those in the original image or video. Therefore, this technology might promote medical video data sharing, improve the implementation of FER systems in clinical settings, and protect sensitive data. Furthermore, deepfake technology will further enhance the potential use of artificial empathy in helping doctors provide empathic care based on patients' emotional experiences to achieve a good doctor-patient therapeutic relationship.

Acknowledgments

This research is funded by the Ministry of Science and Technology (grants MOST 110-2320-B-038-029-MY3, 110-2221-E-038-002-MY2, and 110-2622-E-038-003-CC1).

Conflicts of Interest

None declared.

References

1. Ha JF, Longnecker N. Doctor-patient communication: a review. *Ochsner J* 2010;10(1):38-43 [FREE Full text] [Medline: [21603354](#)]
2. Zachariae R, Pedersen CG, Jensen AB, Ehrnrooth E, Rossen PB, von der Maase H. Association of perceived physician communication style with patient satisfaction, distress, cancer-related self-efficacy, and perceived control over the disease. *Br J Cancer* 2003 Mar 10;88(5):658-665 [FREE Full text] [doi: [10.1038/sj.bjc.6600798](#)] [Medline: [12618870](#)]
3. Sullivan LM, Stein MD, Savetsky JB, Samet JH. The doctor-patient relationship and HIV-infected patients' satisfaction with primary care physicians. *J Gen Intern Med* 2000 Jul;15(7):462-469 [FREE Full text] [doi: [10.1046/j.1525-1497.2000.03359.x](#)] [Medline: [10940132](#)]
4. Renzi C, Abeni D, Picardi A, Agostini E, Melchi CF, Pasquini P, et al. Factors associated with patient satisfaction with care among dermatological outpatients. *Br J Dermatol* 2001 Oct;145(4):617-623. [doi: [10.1046/j.1365-2133.2001.04445.x](#)] [Medline: [11703289](#)]
5. Cánovas L, Carrascosa AJ, García M, Fernández M, Calvo A, Monsalve V, Empathy Study Group. Impact of empathy in the patient-doctor relationship on chronic pain relief and quality of life: A prospective study in Spanish pain clinics. *Pain Med* 2018 Jul 01;19(7):1304-1314. [doi: [10.1093/pm/pnx160](#)] [Medline: [29016846](#)]
6. Ranjan P, Kumari A, Chakrawarty A. How can doctors improve their communication skills? *J Clin Diagn Res* 2015 Mar;9(3):JE01-JE04 [FREE Full text] [doi: [10.7860/JCDR/2015/12072.5712](#)] [Medline: [25954636](#)]
7. Butow P, Hoque E. Using artificial intelligence to analyse and teach communication in healthcare. *Breast* 2020 Apr;50:49-55 [FREE Full text] [doi: [10.1016/j.breast.2020.01.008](#)] [Medline: [32007704](#)]
8. Khan FH, Hanif R, Tabassum R, Qidwai W, Nanji K. Patient attitudes towards physician nonverbal behaviors during consultancy: Result from a developing country. *ISRN Family Med* 2014 Feb 04;2014:473654 [FREE Full text] [doi: [10.1155/2014/473654](#)] [Medline: [24977140](#)]
9. Cuff BMP, Brown SJ, Taylor L, Howat DJ. Empathy: A review of the concept. *Emot Rev* 2014 Dec 01;8(2):144-153. [doi: [10.1177/1754073914558466](#)]
10. Aminololama-Shakeri S, López JE. The doctor-patient relationship with artificial intelligence. *AJR Am J Roentgenol* 2019 Feb;212(2):308-310. [doi: [10.2214/AJR.18.20509](#)] [Medline: [30540210](#)]
11. Iqbal U, Celi LA, Li YCJ. How can artificial intelligence make medicine more preemptive? *J Med Internet Res* 2020 Aug 11;22(8):e17211 [FREE Full text] [doi: [10.2196/17211](#)] [Medline: [32780024](#)]
12. Sebe N, Cohen I, Gevers T, Huang TS. Multimodal approaches for emotion recognition: A survey. 2005 Jan 17 Presented at: Electronic Imaging 2005; January 16-20 2005; San Jose, California, United States. [doi: [10.1117/12.600746](#)]
13. Bänziger T, Grandjean D, Scherer KR. Emotion recognition from expressions in face, voice, and body: the Multimodal Emotion Recognition Test (MERT). *Emotion* 2009 Oct;9(5):691-704. [doi: [10.1037/a0017088](#)] [Medline: [19803591](#)]

14. Lazzeri N, Mazzei D, Greco A, Rotesi A, Lanatà A, De Rossi DE. Can a humanoid face be expressive? A psychophysiological investigation. *Front Bioeng Biotechnol* 2015 May 26;3:64 [FREE Full text] [doi: [10.3389/fbioe.2015.00064](https://doi.org/10.3389/fbioe.2015.00064)] [Medline: [26075199](https://pubmed.ncbi.nlm.nih.gov/26075199/)]
15. Frank MG. Facial expressions. In: Smelser NJ, Baltes PB, editors. *International Encyclopedia of the Social & Behavioral Sciences*. Amsterdam, Netherlands: Elsevier; 2001:5230-5234.
16. Mancini G, Biolcati R, Agnoli S, Andrei F, Trombini E. Recognition of facial emotional expressions among Italian pre-adolescents, and their affective reactions. *Front Psychol* 2018 Aug 03;9:1303 [FREE Full text] [doi: [10.3389/fpsyg.2018.01303](https://doi.org/10.3389/fpsyg.2018.01303)] [Medline: [30123150](https://pubmed.ncbi.nlm.nih.gov/30123150/)]
17. Jack RE, Schyns PG. The human face as a dynamic tool for social communication. *Curr Biol* 2015 Jul 20;25(14):R621-R634 [FREE Full text] [doi: [10.1016/j.cub.2015.05.052](https://doi.org/10.1016/j.cub.2015.05.052)] [Medline: [26196493](https://pubmed.ncbi.nlm.nih.gov/26196493/)]
18. Frith C. Role of facial expressions in social interactions. *Philos Trans R Soc Lond B Biol Sci* 2009 Dec 12;364(1535):3453-3458 [FREE Full text] [doi: [10.1098/rstb.2009.0142](https://doi.org/10.1098/rstb.2009.0142)] [Medline: [19884140](https://pubmed.ncbi.nlm.nih.gov/19884140/)]
19. Saxena A, Khanna A, Gupta D. Emotion recognition and detection methods: A comprehensive survey. *Journal of Artificial Intelligence and Systems* 2020;2:53-79 [FREE Full text] [doi: [10.33969/ais.2020.21005](https://doi.org/10.33969/ais.2020.21005)]
20. Mehendale N. Facial emotion recognition using convolutional neural networks (FERC). *SN Appl Sci* 2020 Feb 18;2(446):1-8 [FREE Full text] [doi: [10.1007/s42452-020-2234-1](https://doi.org/10.1007/s42452-020-2234-1)]
21. Akhand MAH, Roy S, Siddique N, Kamal MAS, Shimamura T. Facial emotion recognition using transfer learning in the deep CNN. *Electronics (Basel)* 2021 Apr 27;10(9):1036 [FREE Full text] [doi: [10.3390/electronics10091036](https://doi.org/10.3390/electronics10091036)]
22. Song Z. Facial expression emotion recognition model integrating philosophy and machine learning theory. *Front Psychol* 2021 Sep 27;12:759485 [FREE Full text] [doi: [10.3389/fpsyg.2021.759485](https://doi.org/10.3389/fpsyg.2021.759485)] [Medline: [34646223](https://pubmed.ncbi.nlm.nih.gov/34646223/)]
23. Uddin MZ. Chapter 26 - A local feature-based facial expression recognition system from depth video. In: Deligiannidis L, Arabnia HR, editors. *Emerging Trends in Image Processing, Computer Vision and Pattern Recognition*. Burlington, Massachusetts: Morgan Kaufmann; 2015:407-419.
24. John A, Abhishek MC, Ajayan AS, Kumar V, Sanoop S, Kumar VR. Real-time facial emotion recognition system with improved preprocessing and feature extraction. 2020 Oct 06 Presented at: 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT); August 20-22, 2020; Tirunelveli, India. [doi: [10.1109/icssit48917.2020.9214207](https://doi.org/10.1109/icssit48917.2020.9214207)]
25. Minaee S, Minaei M, Abdolrashidi A. Deep-emotion: Facial expression recognition using attentional convolutional network. *Sensors (Basel)* 2021 Apr 27;21(9):3046 [FREE Full text] [doi: [10.3390/s21093046](https://doi.org/10.3390/s21093046)] [Medline: [33925371](https://pubmed.ncbi.nlm.nih.gov/33925371/)]
26. Ng HW, Nguyen VD, Vonikakis V, Winkler S. Deep learning for emotion recognition on small datasets using transfer learning. 2015 Nov Presented at: ICMI '15: International Conference on Multimodal Interaction; November 9-13, 2015; Seattle, Washington, USA p. 443-449. [doi: [10.1145/2818346.2830593](https://doi.org/10.1145/2818346.2830593)]
27. Leo M, Carcagni P, Mazzeo PL, Spagnolo P, Cazzato D, Distante C. Analysis of facial information for healthcare applications: A survey on computer vision-based approaches. *Information (Basel)* 2020 Feb 26;11(3):128 [FREE Full text] [doi: [10.3390/info11030128](https://doi.org/10.3390/info11030128)]
28. Surcinelli P, Andrei F, Montebanocci O, Grandi S. Emotion recognition of facial expressions presented in profile. *Psychol Rep* 2021 May 26;332941211018403. [doi: [10.1177/00332941211018403](https://doi.org/10.1177/00332941211018403)] [Medline: [34039106](https://pubmed.ncbi.nlm.nih.gov/34039106/)]
29. Nguyen TT, Nguyen QVH, Nguyen DT, Nguyen DT, Huynh-The T, Nahavandi S, et al. Deep learning for deepfakes creation and detection: A survey. *arXiv Preprint* posted online on February 6, 2022 [FREE Full text]
30. Guarnera L, Giudice O, Battiato S. DeepFake detection by analyzing convolutional traces. 2020 Jul 28 Presented at: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); June 14-19, 2020; Seattle, Washington, USA. [doi: [10.1109/cvprw50498.2020.00341](https://doi.org/10.1109/cvprw50498.2020.00341)]
31. Hosler B, Salvi D, Murray A, Antonacci F, Bestagini P, Tubaro S, et al. Do deepfakes feel emotions? A semantic approach to detecting deepfakes via emotional inconsistencies. 2021 Sep 01 Presented at: 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); June 19-25, 2021; Nashville, Tennessee, USA p. 1013-1022.
32. Zhu B, Fang H, Sui Y, Li L. Deepfakes for medical video de-identification: Privacy protection and diagnostic information preservation. 2020 Feb Presented at: AIES '20: AAAI/ACM Conference on AI, Ethics, and Society; February 7-9, 2020; New York, New York, USA p. 414-420. [doi: [10.1145/3375627.3375849](https://doi.org/10.1145/3375627.3375849)]
33. Floridi L. Artificial intelligence, deepfakes and a future of ectypes. *Philos Technol* 2018 Aug 1;31(3):317-321. [doi: [10.1007/s13347-018-0325-3](https://doi.org/10.1007/s13347-018-0325-3)]
34. Kietzmann J, Lee LW, McCarthy IP, Kietzmann TC. Deepfakes: Trick or treat? *Bus Horiz* 2020;63(2):135-146. [doi: [10.1016/j.bushor.2019.11.006](https://doi.org/10.1016/j.bushor.2019.11.006)]
35. Zhizhong G, Meiying S, Yiyu S, Jiahua W, Ziyun P, Minghong Z, et al. Smart Monitoring Re-Upgraded--Introducing Facial Emotion Recognition [in Chinese]. *Computer and Communication* 2018 Oct 25(175):29-31. [doi: [10.29917/CCLTJ](https://doi.org/10.29917/CCLTJ)]
36. Siarohin A, Lathuilière S, Tulyakov S, Ricci E, Sebe N. First order motion model for image animation. 2019 Dec Presented at: 33rd International Conference on Neural Information Processing Systems; December 8-14, 2019; Vancouver, Canada p. 7137-7147.
37. Markides M. The importance of good communication between patient and health professionals. *J Pediatr Hematol Oncol* 2011 Oct;33 Suppl 2:S123-S125. [doi: [10.1097/MPH.0b013e318230e1e5](https://doi.org/10.1097/MPH.0b013e318230e1e5)] [Medline: [21952568](https://pubmed.ncbi.nlm.nih.gov/21952568/)]

38. Kerasidou A. Artificial intelligence and the ongoing need for empathy, compassion and trust in healthcare. *Bull World Health Organ* 2020 Apr 01;98(4):245-250 [[FREE Full text](#)] [doi: [10.2471/BLT.19.237198](https://doi.org/10.2471/BLT.19.237198)] [Medline: [32284647](https://pubmed.ncbi.nlm.nih.gov/32284647/)]
39. Thom DH, Hall MA, Pawlson LG. Measuring patients' trust in physicians when assessing quality of care. *Health Aff (Millwood)* 2004;23(4):124-132. [doi: [10.1377/hlthaff.23.4.124](https://doi.org/10.1377/hlthaff.23.4.124)] [Medline: [15318572](https://pubmed.ncbi.nlm.nih.gov/15318572/)]
40. Bogiatzaki V, Frengidou E, Savakis E, Trigoni M, Galanis P, Anagnostopoulos F. Empathy and burnout of healthcare professionals in public hospitals of Greece. *Int J Caring Sci* 2019;12(2):611-626 [[FREE Full text](#)]
41. Jeong YU, Yoo S, Kim YH, Shim WH. De-identification of facial features in magnetic resonance images: Software development using deep learning technology. *J Med Internet Res* 2020 Dec 10;22(12):e22739 [[FREE Full text](#)] [doi: [10.2196/22739](https://doi.org/10.2196/22739)] [Medline: [33208302](https://pubmed.ncbi.nlm.nih.gov/33208302/)]

Abbreviations

AI: artificial intelligence

FER: facial emotion recognition

GAN: generative adversarial network

Edited by A Mavragani; submitted 13.04.21; peer-reviewed by S Bidmon, A Hashim; comments to author 11.08.21; revised version received 06.12.21; accepted 28.12.21; published 04.03.22.

Please cite as:

Yang HC, Rahmanti AR, Huang CW, Li YCJ

How Can Research on Artificial Empathy Be Enhanced by Applying Deepfakes?

J Med Internet Res 2022;24(3):e29506

URL: <https://www.jmir.org/2022/3/e29506>

doi: [10.2196/29506](https://doi.org/10.2196/29506)

PMID: [35254278](https://pubmed.ncbi.nlm.nih.gov/35254278/)

©Hsuan-Chia Yang, Annisa Ristya Rahmanti, Chih-Wei Huang, Yu-Chuan Jack Li. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 04.03.2022. 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.

Viewpoint

Tackling the Burden of Electronic Health Record Use Among Physicians in a Mental Health Setting: Physician Engagement Strategy

Tania Tajirian^{1,2}, MD; Damian Jankowicz¹, PhD; Brian Lo^{1,3}, BHSc, MHI; Lydia Sequeira^{1,3}, MHI, PhD; Gillian Strudwick^{1,3,4}, RN, PhD; Khaled Almilaji¹, MHI, MD; Vicky Stergiopoulos^{5,6}, MHSc, MD

¹Information Management Group, Centre for Addiction and Mental Health, Toronto, ON, Canada

²Department of Family and Community Medicine, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

³Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada

⁴Centre for Complex Interventions, Centre for Addiction and Mental Health, Toronto, ON, Canada

⁵Physician-in-Chief Office, Centre for Addiction and Mental Health, Toronto, ON, Canada

⁶Department of Psychiatry, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:

Tania Tajirian, MD

Information Management Group

Centre for Addiction and Mental Health

100 Stokes Street

Toronto, ON, M6J 1H4

Canada

Phone: 1 416 535 8501 ext 30515

Email: Tania.Tajirian@camh.ca

Abstract

The burden associated with using the electronic health record system continues to be a critical issue for physicians and is potentially contributing to physician burnout. At a large academic mental health hospital in Canada, we recently implemented a Physician Engagement Strategy focused on reducing the burden of electronic health record use through close collaboration with clinical leadership, information technology leadership, and physicians. Built on extensive stakeholder consultation, this strategy highlights initiatives that we have implemented (or will be implementing in the near future) under four components: engage, inspire, change, and measure. In this viewpoint paper, we share our process of developing and implementing the Physician Engagement Strategy and discuss the lessons learned and implications of this work.

(*J Med Internet Res* 2022;24(3):e32800) doi:[10.2196/32800](https://doi.org/10.2196/32800)

KEYWORDS

burnout; organizational strategy; electronic health record use; clinical informatics; medical informatics

Introduction

Background

With growing levels of clinician burnout both before and during the COVID-19 pandemic, the burden associated with the use of electronic health record (EHR) systems has emerged as a paramount challenge [1]. In particular, with an increase in reporting and clinical research needs required by consumers of health data (ie, administrators and researchers) during the spread of the pandemic, there have been additional burden for data generators (ie, providers who document care in the EHR) [2]. This has resulted in clinicians spending more time using the EHR system to complete documentation than actual patient

care. The recent call to action by Shanafelt [3] highlighted the importance of thinking about how digital technologies are being introduced as another component into the clinical environment. In particular, there is a greater demand to think about better designed solutions that fit the needs of clinicians [3]. In addition to reports by physicians on the impact of EHR burden on burnout [4,5], this has resulted in numerous recommendations from organizations such as the Ontario Medical Association [6] on reconsidering the impact and use of technology in clinical practice. Therefore, it is likely that EHR-related burden, in addition to the efforts and challenges involved in managing the pandemic, have collectively led clinician burnout to an all-time high [7,8].

As EHR systems are increasingly enhanced with advanced features to improve patient care, such as clinical decision support and predictive analytics, the impact of these capabilities on clinicians' ability to effectively and meaningfully deliver care must not be forgotten. As highlighted in many commentaries on this topic and most recently, in a chapter on EHR burden by the National Academies of Sciences, Engineering, and Medicine [9], the use of the EHR must not detract from the core aspects of medicine, such as the therapeutic relationship, nor cause unnecessary frustration, complexity, and burden on clinicians [10]. In this context, many initiatives in this field have recently emerged. For example, the *Clickbusters* initiative, developed by the Vanderbilt Clinical Informatics Centre [11], aims to reduce the number of unnecessary alerts delivered to end users, thereby reducing EHR burden and burnout.

Of note, most efforts to date that address EHR-related burden for clinicians have focused on the US context, with scant evidence emerging from other countries. Given that each country has varying practice and clinical requirements, the factors and bottlenecks that influence EHR burden in various settings are likely different. For example, in a recent survey on EHR burnout conducted as part of the 25x5 Symposium [12] in the United States, documentation requirements for reimbursement were cited as a main factor leading to EHR burden for physicians. These challenges are specific to the US context and do not apply to many other countries, including Canada, where health care is publicly funded. However, even without such reimbursement-driven documentation requirements in Canada, we face substantial EHR-related burnout rates [13] comparable with those in other countries. In addition, it is expected that the challenges associated with EHR burden differ across disciplines. For example, given that mental health relies heavily on narrative documentation, the pain points in using the EHR would likely differ from those experienced by specialties that rely on the use of forms or structured templates. Without sufficient discussion and evidence in the literature on how mental health care organizations in Canada are managing these challenges, there is a lack of guidance for these organizations to build relevant and effective initiatives in their own settings. Moreover, it is unclear how existing best practices for system development and implementation (eg, super users) should be best leveraged to address EHR burden. Given these evidence gaps, we seek to contribute to the emerging body of evidence and support the collective effort of reducing EHR burden for all disciplines across countries.

In this viewpoint paper, we describe our efforts to reduce EHR burden for physicians at a large academic mental health hospital in Canada. Building on our previously published studies focused on needs assessment, implementation, and evaluation of individual initiatives [4,14,15], this viewpoint shares the development and implementation of our overarching strategy. We discuss our Physician Engagement Strategy, which aims at identifying and addressing opportunities for EHR improvements at our site. On the basis of our experience to date, we conclude with key success factors and lessons learned in developing and implementing the initiatives included in this strategy.

The Organization

The Physician Engagement Strategy was implemented in a large academic mental health hospital that provides care to >34,000 patients experiencing mental illness in Toronto, Ontario, Canada. In an effort to improve the quality and continuity of patient care, the organization implemented an integrated EHR system (Cerner Millennium) in 2014. Now, the organization offers a paperless care environment, where all processes, from orders to medication administration, are conducted through and with the EHR system. With a *single source of truth*, the organization has since been able to improve key quality of care outcomes by enhancing medication safety [16] and embedding psychiatric risk flags in the system [17]. This has resulted in obtaining the highest rating (stage 7) on the Electronic Medical Record Adoption Model [18] and level 6 on the Adoption Model for Analytics Maturity [19] from the Healthcare Information and Management Systems Society. Most notably, the organization was awarded the prestigious Davies Enterprise Award in 2018 [20]. On average, the EHR system is used by >400 physicians to deliver care at the organization.

Despite these achievements and the organization's focus on improving the safety and quality of patient care, physicians have had mounting concerns regarding the usability of the EHR in their daily workflows, highlighting it as a major source of burnout in our 2017 physician wellness survey (conducted with all physicians across the hospital) [15]. Our organization's wider Physician Engagement, Wellness, and Excellence Strategy included several interventions to improve physician support at the individual, team, and organizational levels. One of the six proposed initiatives under this strategy included *efforts to optimize use of EHRs to enhance the efficiency of practice* [15]. In alignment with departmental leadership (VS) needs to optimize the use of EHR and reduce the associated burden for clinicians, leadership (DJ) from information management prioritized clinician-driven innovation, including efforts to address EHR-related clinician burnout. Consequently, in 2018, the inaugural Chief Medical Information Officer (CMIO) was tasked with improving physicians' experiences with the EHR to reduce EHR-related burden. Given the importance of leadership buy-in and prioritization of strategies, the CMIO developed and implemented a multipronged strategy to address the ongoing and emerging challenges for physicians related to the use of the EHR.

Building Our Physician Engagement Strategy

As part of the formative work toward developing and implementing a tailored Physician Engagement Strategy, we undertook a needs assessment in 2019 to understand the main challenges experienced by and EHR-related goals of each academic division within our organization and reviewed the literature to identify strategies and initiatives that could help to address these challenges and aspirations. This effort and the resulting strategy are further described below.

Needs Assessment

Overview

As part of preparatory work to inform and align initiatives targeting EHR burden on physicians to the needs of the organization, it was essential to achieve a holistic understanding of the frustrations and challenges that hinder the effective and meaningful use of the EHR across our organization. A tour of the academic divisions was undertaken and a benchmark survey was launched in 2019 to collect direct feedback and insights on burnout and the current state of EHR use by physicians at our organization. The divisional tour focused on obtaining feedback mostly from physician leadership across the organization, whereas the benchmark survey more broadly and anonymously captured the voice of frontline physicians. This ensured that a balanced *bottom-up* leadership approach was embedded in the development of the strategy. It should be noted that this divisional tour was conducted when several foundational articles (eg, *Why doctors hate their computer* by Gawande [21]) appeared in the published literature highlighting the issues for the medical community, including within the Canadian context [22]. As such, there was great interest from both the organization and frontline physicians to participate in these activities.

Divisional Tour

Through the divisional tour, our hospital's CMIO visited each of the 7 academic physician divisions' monthly meetings and gathered feedback on their top 3 priorities for EHR optimization. There were 2 main purposes for the divisional tour. First, given that the benchmark survey focused on collecting individual feedback from frontline physicians, the divisional tour helped to contextualize the results through an in-depth discussion with the team. Second, it was also used to obtain buy-in from frontline and clinician leaders. Buy-in was a critical part of the success of these initiatives, and this opportunity provided another forum for physicians to contribute their ideas and perspectives.

In addition to discussions with clinical leadership during these meetings, frontline physicians were also consulted through the existing team huddles for each academic division to obtain their perspective on these issues. It should be noted that this divisional tour was conducted 5 years after the introduction of the hospital's EHR system. From these consultations, the identified improvement priorities included documentation, orders, and chart navigation through the EHR. More specifically, within *documentation*, there were requests to improve standardized templates and auto-population, increase access to speech recognition technology, and reduce clicks. Second, regarding *orders*, physicians identified the need to increase practice-specific order sets, reduce clicks within workflows, and simplify order measures. Third, in *chart navigation*, there were requests to make the EHR more user-friendly and intuitive with respect to finding relevant information. The automatic inclusion of all laboratory test results and previous medications through the EHR has led to an increase in note sizes that contain multiple pages of nonessential information, which is a phenomenon termed as *note bloat* [23,24], and is frequently identified as a concern for patient safety. Other priorities were also identified, such as the desire to integrate physician billings

within the EHR and to simplify the process of discharge medication reconciliation.

Benchmark Survey

Following this tour, we conducted a benchmark survey to identify the extent of burnout among our physicians and identify the significant EHR-related contributors to physician burnout [4]. The methods and approach are detailed in a separate paper [4]. We achieved a high survey response rate among full-time physicians (156/208, 75%), with 25.6% (45/176) of physicians reporting having one or more symptoms of burnout. Furthermore, 74.5% (155/208) of respondents who reported burnout symptoms identified the EHR as a contributor. Safety concerns with the EHR and efficient communication were 2 factors that differentiated the groups that were burned out from those that were not. Respondents who had high satisfaction with the EHR used work-arounds to complete tasks (eg, copy and paste from word processing software) or were super users with knowledge of documentation *shortcuts*. Those with low EHR-satisfaction reported excessive clicks and time sinks with using the EHR [4].

Overall, our survey demonstrated that there was a critical need to mitigate EHR-related burden and the associated burnout by optimizing the EHR to fit within physicians' workflows and by improving awareness of EHR best practices.

Literature Review

Consistent with academic practices, we sought to inform our Physician Engagement Strategy with the latest evidence. Given the number of interventions aimed at combating EHR-related burnout in the published literature, our team conducted a rapid literature review on this topic [25]. From a review of 50 related articles published between 2014 and 2019, we found that the measurement of EHR burden needs to be performed both subjectively (eg, via surveys) and objectively (eg, using use data) [26]. We also identified that interventions to reduce EHR burnout were focused on four main aspects: enhancing and redesigning the interface of EHR screens, delivering tailored education and training to end users, improving communication, and providing additional support for administrative tasks [25]. To implement many of these interventions, approaches similar to a Plan-Do-Study-Act cycle for quality improvement [27] were used, which was in alignment with our needs assessment process (ie, measuring *pain points*). Through our literature review, we also learned that following the implementation of interventions, it is essential to continue enhancing and optimizing these through further evaluation using analytics data, surveys, and other methods. Although findings from our literature review have laid the foundational building blocks of our strategy, it was important to adapt and align these strategies and interventions to the specific challenges present within our organization.

The Physician Engagement Strategy

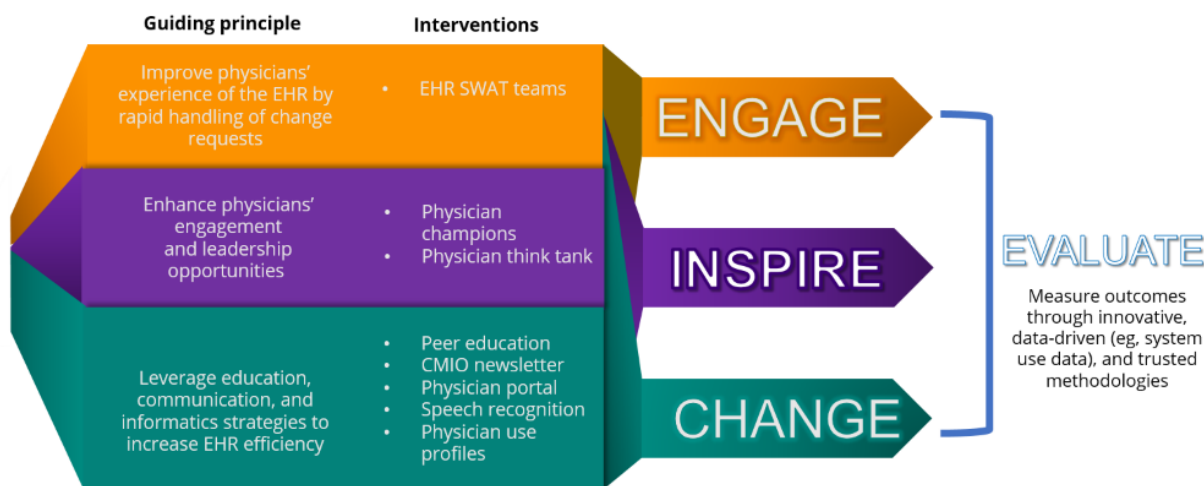
Overview

On the basis of our in-depth needs assessment and review of the literature, we developed our first iteration of the Physician Engagement Strategy in 2020. In the development of this

physician-centric multipronged approach, we had in-depth discussions with our information technology (IT) and clinical leadership teams to ensure feasibility and availability of resources. An overview of the strategy is depicted in Figure 1. The strategy comprises four main goals: (1) improving the handling and resolution of EHR issues; (2) enhancing physician

engagement and leadership opportunities on EHR-related decisions; (3) leveraging communication, education, and informatics strategies to increase efficient and meaningful use of the EHR; and (4) measuring the impact of these strategies to achieve data-driven insights.

Figure 1. Physician Engagement Strategy. CMIO: Chief Medical Information Officer; EHR: electronic health record.



Throughout the implementation of initiatives to support these goals, 3 main guiding principles (pillars) were considered essential for success. Foremost, it was critical to *engage* all relevant stakeholders and frontline physicians to ensure that their perspectives are heard and any challenges are brought up for examination. We also considered it important to provide physicians an opportunity to be *inspired* to participate in leadership roles and be involved in the decision-making process of EHR changes that impact patient care and physicians' use of the EHR. Finally, we detailed a multitude of initiatives aimed at collectively supporting the ability to *change* the use of the EHR system such that it improves efficiency and end user experience. Throughout these pillars, we also ensured that there was a focus on measuring outcomes to *evaluate* whether the specific initiatives have worked as intended. The initiatives outlined in the strategy are described in detail below.

Engage: Improve Physician’s Experience With the EHR by Rapid Handling of Change Requests (SWAT Teams)

To engage physicians across our organization, we developed an EHR SWAT team initiative through adaptation of initiatives [28] identified from our literature review. Traditional governance models have focused on identifying and implementing requirements solely by the IT team [3]. However, from the IT perspective, there is often a lack of understanding of the actual requirements needed for the change to be impactful. Moreover, many of the changes often affect other clinical areas (eg, pharmacy and laboratory). Thus, implementing in isolation can lead to more downstream challenges. As such, the SWAT team model, which mirrors other initiatives at the University of Colorado School of Medicine [28,29], overcomes this challenge by bringing together a collaborative team to discuss and identify a commonly agreed set of requirements for each issue. Our team-based intervention (SWAT) included assembling

an interdisciplinary team of specialists including our CMIO; clinical informatics nurses and educators; and representatives from pharmacy informatics, health information management, clinical applications, and project management [14]. Through this intervention, we met with physicians from each of the seven academic divisions across our organization, collected EHR change requests, and prioritized them into four categories: (1) additional education, (2) quick fixes (<6 weeks), (3) future fixes (≥1 year), and (4) unable to address owing to technical or regulatory restraints. In total, we gathered 118 requests (eg, including adding keyword search functionality, minimizing freezing, and auto-faxing) [14].

Inspire: Enhance Physician’s Engagement and Leadership Opportunities

Physician Champions

As part of our focus on inspiring and fostering physicians’ voices in the decision-making process, we designated physician EHR champions (*liaisons*) tasked with liaising with all physicians in their division and bringing forward pain points and recommended changes on an ongoing basis. Divisional liaisons were nominated annually by both their peers and divisional leadership and became the link between different stakeholders, helping us make meaningful EHR changes. These individuals were key players within our EHR SWAT initiative. Liaisons’ responsibilities continued to evolve with the changing needs of the initiatives, and they became the pilot user group for future technology and informatics applications.

Physician Think Tanks

Discussions with physician champions highlighted the need for a cross-divisional lens to identify the applicability of EHR changes. As such, a venue to address this gap and facilitate discussion between physician divisional liaisons and other relevant stakeholders was needed. We implemented monthly

Physician Think Tanks in 2019, which focused on the successful use of the EHR to improve patient safety and quality of care. These meetings are chaired by our CMIO and are attended by physician champions from each academic division and relevant clinical (eg, pharmacy, laboratory, and diagnostics informatics and professional practice) and IT leaders (eg, clinical applications), with composition similar to that of our SWAT team. Before each of these meetings, stakeholders are encouraged to bring forward challenges and questions for discussion at these meetings.

Appropriate engagement with stakeholders is critical for implementing digital tools in a meaningful manner that aligns with the needs of end users, and therefore, we used this forum before the implementation of any new initiatives. As these meetings evolved since their inception in 2019, the venue became a versatile space for EHR optimization. New ideas and features (eg, optimizing order set and reducing auto-population content) are demonstrated at these meetings to collect initial feedback from clinicians. Updates on initiatives (eg, SWAT) and implementations are also presented at these meetings to help support brainstorming of the feasibility and availability of resources. From a quality improvement perspective, evaluation results are presented to solicit feedback from the group for contextualizing the results and identifying next steps for optimization of the EHR.

Change: Leverage Communication, Education, and Informatics Initiatives to Increase EHR Efficiency

Communication: CMIO Monthly Newsletter

One of the main challenges identified from our benchmark survey was the lack of appropriate communication channels for EHR-related updates (eg, policy and technical changes). Although the organization provided EHR updates to all users on a regular basis, these email updates were incomplete and not tailored to physicians. Given that these EHR changes usually coincide with broader organizational policy changes and mandates, physicians reported a lack of a single source of truth for any EHR-related updates. These issues can cause confusion and inconsistency in EHR use across the organization, leading to concerns regarding patient safety and quality of care.

To address this issue, the CMIO monthly newsletter was developed. These newsletters are developed with in-depth consultation with clinical (eg, clinical informatics nurse educators) and IT stakeholders to ensure that the content is relevant and useful. The content of the newsletter varies each month and depends on the changes and discussions at that time (Figure 2). Examples of content include initiative updates, interviews with digital health leaders, EHR tips and tricks, related literature, IT changes and EHR changes for health records, pharmacy and therapeutics, and laboratory and diagnostics related EHR updates and clarifications. The newsletter was also a critical method to close the loop of communication on all changes requested through our SWAT initiative. The monthly read rates of these newsletters continue to increase, with our last newsletter (ie, April 2021) being opened by >208 physicians (approximately 50% open rate).

Figure 2. Chief Medical Information Officer monthly newsletter (table of contents). FAQ: frequently asked questions.

TANIA'S CORNER	NEW	Peer Education – Discharge Summary & How To Autofax	1	
	NEW	Clinician Recruitment (Mental Health Notes)	2	
NEWSFLASH	NEW	NEW I-CARE Login Process	3	
	NEW	Additional Drive for “Clinical Access” Users	4	
	UPDATE	Revised Instructions for Prescription Faxing	5	
	Q&A	A Quick FAQ for Physicians	6	
	UPDATE	How To Obtain Secure Mail on Your Phone	7	
	TIP OF THE MONTH	Manage Proposed Orders and Order Sets (Supervisors)	8	
	I-CARE UPDATES	CHANGE	Sending a Private Lab Requisition	9
		HEALTH RECORDS	REMINDER: How to find COVID Vaccine Information	10

Communication: Physician Portal

In addition to the issues related to communication, another intervention was developed to support navigation and rapid finding of EHR-related information. The *Physician Portal* is a collaborative initiative with the physician-in-chief to provide physicians with a one-stop web-based location for all information relevant to physicians, such as wellness initiative updates, policies, and educational presentations. Within the portal's EHR tips and tricks section, all EHR-related information can be found. Previous CMIO monthly newsletters are made available through this portal, and physicians are able to use a search functionality to find relevant information across the site. Our peer education videos are also hosted on this portal. Recently, tip sheets for EHR use have been added to the site, and physicians are encouraged to visit the site to access these resources. The centralized location for EHR-related information is expected to reduce the navigation burden for EHR-related information.

Education: Peer Education Videos

Electronic learning modules have become a staple for providing training and support for a wide range of practice- and policy-related topics. However, their uptake and effectiveness for EHR training and best practices for physicians remain very limited. Given that this user group is inundated with information daily, it was essential to develop targeted messaging that aligns with their needs and questions [30]. As such, peer education was identified to be a useful approach for addressing these issues. Peer education through super users can support end users in mastering the use of the EHR, and previous peer-led EHR training initiatives in Southern California Permanente Medical Group have found that it can save physicians 4 to 5 minutes per hour, equating to approximately 40 minutes a day [31].

Although peer EHR education is typically done in real time and in person, social distancing restrictions of the COVID-19 pandemic have made this difficult to achieve. As a result, we piloted the development of peer education videos. These peer education videos are short in length (3-5 minutes) and focus on specific knowledge gaps found across the organization. During these videos, super users are invited to provide education and guidance (eg, demonstration of a workflow) to the audience based on their experiences and best practices. These videos are posted on our *Physician Portal*, and communication to increase awareness of these tools are done through the monthly CMIO newsletter and the weekly departmental physician newsletter. Physicians are invited to watch the videos in their own time and pace. Currently, two modules have been developed (medication reconciliation and discharge summaries) and a few other videos are in development.

Informatics: Speech Recognition Technology

Documentation burden was identified as an important issue for physicians in both the benchmark survey and the divisional tour. In particular, given that psychiatric documentation is fairly narrative in nature, physicians spend significant amount of time typing directly into their EHR system. As a result, documentation methods remain a significant pain point for the organization.

Speech recognition technology has been identified as a suitable solution for mitigating documentation burden in the EHR. Evaluations conducted in other settings have found that physicians report satisfaction with speech recognition technology and its utility in reducing the burden. At our organization, an older version of speech recognition software was procured and deployed to a small number of physicians [32]. However, the limited licenses and lack of a concerted support strategy led to its suboptimal adoption across the organization. Given the renewed focus on this issue, we endeavored to roll out an improved version of speech recognition technology that features improved accuracy, a mobile app microphone, and cloud-based dictation engine. As part of this rollout, all physicians and residents will have access to speech recognition technology for documentation in the EHR system [30].

The speech recognition service will be delivered collaboratively with the other initiatives of the strategy in several ways. The EHR SWAT team intervention will be used to collect technical issues and feature improvements with regard to the platform. The Physician Portal and CMIO newsletters will be used to communicate improvements to the platform and encourage physicians to receive optimization training and education using the platform.

Informatics: Physician EHR Use Profiles

Another complementary intervention to improve physicians' awareness of their EHR practice and deliver feedback was the use of dashboards. Direct feedback dashboards have been used widely across medicine for outcomes such as improving compliance for venous thromboembolism prophylaxis [33], reducing imaging use [34], and improving pediatric emergency care [35]. Dashboards convey information through the use of visual representation of data to help amplify cognition and capitalize on human perceptual capabilities [36]. Our *physician use profiles* (Figure 3) are dashboards that display information on an individual physician level, allowing physicians to view their own EHR system use including metrics such as time spent within the EHR per patient, time spent in documentation, percentage of after-hours use, number of clicks per order, and other measures of efficiency. Through these dashboards, physicians can also compare their EHR use with their divisional average and identify whether they need to lean into the various initiatives of the Physician Engagement Strategy, such as additional training through EHR SWAT meetings, peer education videos, or speech recognition technology. We also anticipate physicians to self-identify as super users through the use of these dashboards (as only the individual physician can look at their own data) and contribute to peer education within their division. Physicians can view their metrics for a specific period (eg, before and after using speech recognition), with provided context as to how each of the metrics is being calculated. Our team is currently in the process of finalizing the design and content of these *physician EHR use profiles* through extensive consultation with the physician divisional liaisons, academic chiefs, and Physician-In-Chief to make it a meaningful and useful intervention.

Figure 3. Physician electronic health record use profiles.

Evaluate: Measure Outcomes Through Innovative and Trusted Methodologies

Overview

Through our many complementary and interconnected initiatives, we have ensured that time, effort, and resources were used to evaluate each of them. Throughout the development of each intervention, we embedded an evaluation approach to determine whether we have achieved the objectives of the intervention and to identify approaches to streamline or optimize the intervention. This aligns closely with best practices [37] and the Plan-Do-Study-Act cycle [38] of quality improvement. The sources and examples of evaluation initiatives are elaborated in greater detail below.

System Use Data

The role of back-end use data has become an important source of information for identifying user-specific challenges within the EHR system and for guiding tailored and customized training. Within our organization, the use of *EHR use metrics* allow us to objectively measure the impact of our initiatives. Our EHR vendor's back-end analytics platforms (ie, Cerner *Lights On* and Cerner *Advance* [39]) provide detailed analyses of EHR use for all users including physicians, nurses, and those in other disciplines. Leveraging key metrics such as total time spent per patient in documentation, chart review, ordering, and medication reconciliation of the electronic medical record has allowed us to identify whether a specific intervention (eg, speech recognition technology) has impacted the targeted metric (ie, documentation time). There are several other metrics that can be explored, such as contextual metrics (eg, number of patients that a physician has seen over the past month and the percentage of time physicians spend in the EHR after hours) and workflow

metrics (eg, number of clicks used per order and use of order set). Following a thorough validation of how the data within this system compares with our physicians' actual use, we are slowly beginning to leverage this tool to help learn about our physicians' experience with the EHR, eventually working to reduce EHR burden.

Similarly, we are also leveraging the use of *system use metrics* from other technologies that we have implemented, such as our speech recognition solution. For example, personalized messages are periodically delivered to physicians who have access to speech recognition technology and are not using the platform, to identify if they require any technical support or additional training.

Surveys, Interviews, and Focus Groups

When system use data collection is not feasible or fit for evaluation, we conducted evaluations through surveys, interviews, and focus groups, leveraging specific channels of our strategy such as the physician divisional liaisons where possible. For our EHR SWAT initiative, we used short anonymous surveys (with 5-point Likert scales and free-text questions) following the divisional meeting to gather physicians' satisfaction with the initiative [14]. In all, 61% (28/46) of the physicians reported that the intervention increased their proficiency in using EHR functionalities [14]. Surveys and interviews were also leveraged to measure the implementation feasibility of speech recognition, complementary to documentation time outcomes measured from system use data.

Discussion

Overview

To our knowledge, the Physician Engagement Strategy is one-of-a-kind collaborative approach in a Canadian mental health setting that aims to engage both frontline physicians, physician leadership, and IT leaders in reducing EHR burden. Previous national strategies in the United States to reduce clinician burnout have included recommendations such as engaging clinicians in the design and deployment of health IT to ensure the effectiveness, efficiency, usability, and safety of the technology [9], which we have implemented thoroughly throughout our strategy, especially within the *engage* and *inspire* pillars. The Office of the National Coordinator for Health Information Technology in the United States has also released its Strategy on Reducing Regulatory and Administrative Burden Relating to the Use of Health IT and EHRs, which includes recommendations such as better alignment of the EHR system design with real-world clinical workflows, increase in end user training, and improvement of the clinical documentation functionality [40], which we implemented through our EHR SWAT teams, peer education, and speech recognition initiatives, respectively. Our development of a shared vision and approach enables a concerted strategy developed and implemented with feedback and alignment across all stakeholder groups. As the digital health ecosystem (eg, web-based care) continues to become integral to clinical care, the numerous venues for discussion across multiple departments allow for unbiased feedback and opportunities to align the road maps of the EHR and related technologies.

Lessons Learned

As we enter the third year of implementing the EHR Physician Engagement Strategy, a few *lessons learned* that will guide future approaches to optimize the reach and impact of our strategy were identified. Foremost, it is essential that frontline physicians are recognized as the main stakeholders and decision makers in our strategy. Across all our initiatives, careful deliberation and stepwise iterative approaches were used to embed physicians' perspectives and feedback in the implementation and to roll out plans. These stepwise approaches ensured that any *red flags* can be considered and mitigated before the roll out of any initiatives. For example, for speech recognition technology, extensive consultation sessions were conducted before the development of the implementation plan. Understanding physicians' desires and needs helped curate the main components to consider in such an implementation plan. In addition, these discussions should be considered in multidisciplinary forums. Given that the organization has >70 clinics, each specializing in different diagnoses, treatment, or patient populations, it is expected that significant heterogeneity exists in terms of workflows and EHR use patterns. Ensuring that IT and clinical stakeholders from these clinics are engaged in the decision-making process reduces the possibility of unexpected roadblocks or unintended consequences during

implementation. In addition, we realized that periodic review of the initiatives can yield synergistic opportunities to better align and reinforce initiatives within the strategy. In particular, interconnected initiatives can inform each other and maximize their success. For example, during the implementation of speech recognition, the Physician Think Tank was used to identify physicians' needs before the implementation. In addition, the CMIO newsletter was found to be a useful way to communicate updates and success outcomes to all physicians at the organization. Thus, the initiatives of this strategy enable a concerted effort to reduce EHR burden.

Future Directions

As we continue to implement and expand the Physician Engagement Strategy, it has become increasingly important to consider this work in the broader field of EHR burden and digital technologies. We expect that this road map will continue to expand as new features and technologies are becoming embedded within the organization. In this accord, we highlight a few important next steps that we hope to achieve in the coming years. First, many of the interventions available in the literature have not focused on evaluating their impact on EHR burden. With the advent of EHR back-end use data, there is a timely opportunity to evaluate the impact of these initiatives on efficiency and satisfaction with using the EHR. In the next year, we will focus on evaluating and determining the impact of these initiatives using our established evaluation methodologies. We will also explore the perceptions and experiences of clinicians in participating in these initiatives (eg, multidisciplinary groups and super users) to address EHR-related burden [37]. In addition, a number of emerging trends, such as measurement-based care [41,42] and web-based care [43], are becoming evident in our digital mental health infrastructure. As these models of care are streamlined, it would be useful to evaluate their impact on EHR burden. Future work should explore how these initiatives can be made useful prospectively in the planning and implementation of these digital models of care at the organizational level. Finally, we are currently exploring opportunities to embed these initiatives as part of creative professional activities for academic promotions. Doing so may help to support the development of *hybrid physicians* [44] who are equipped to enable the effective use of informatics by physicians to deliver better mental health care.

Conclusions

As EHR-related burden continues to be a critical challenge for many health care systems, we introduce our EHR Physician Engagement Strategy as an approach to reducing these unintended consequences. This strategy involves a multipronged approach that aims to engage clinicians in reimagining their future EHR experiences and empower them as central stakeholders and advocates for digital technologies to achieve the quadruple aim of health care. Future work should focus on evaluating the impact of these initiatives on EHR burden and expanding the impact of this work to other digital health tools at the organization.

Acknowledgments

This strategy is supported by the Centre for Addiction and Mental Health. The authors would like to acknowledge the wonderful support from all physicians at the organization for this work, in particular, the Think Tank and SWAT team members. In addition, the authors would like to thank all clinical and information technology leaders at the organization for their time and support in contributing to the initiatives of this strategy. Finally, the authors would like to thank Sonia Pagliaroli and her team from Cerner Canada for her support with the Lights On platform.

Conflicts of Interest

None declared.

References

1. Bakken S. Can informatics innovation help mitigate clinician burnout? *J Am Med Inform Assoc* 2019;26(2):93-94 [FREE Full text] [doi: [10.1093/jamia/ocy186](https://doi.org/10.1093/jamia/ocy186)] [Medline: [30668747](https://pubmed.ncbi.nlm.nih.gov/30668747/)]
2. Sudat SE, Robinson SC, Mudiganti S, Mani A, Pressman AR. Mind the clinical-analytic gap: electronic health records and COVID-19 pandemic response. *J Biomed Inform* 2021;116:103715 [FREE Full text] [doi: [10.1016/j.jbi.2021.103715](https://doi.org/10.1016/j.jbi.2021.103715)] [Medline: [33610878](https://pubmed.ncbi.nlm.nih.gov/33610878/)]
3. Shanafelt TD. Physician well-being 2.0: where are we and where are we going? *Mayo Clin Proc* 2021;96(10):2682-2693 [FREE Full text] [doi: [10.1016/j.mayocp.2021.06.005](https://doi.org/10.1016/j.mayocp.2021.06.005)] [Medline: [34607637](https://pubmed.ncbi.nlm.nih.gov/34607637/)]
4. Tajirian T, Stergiopoulos V, Strudwick G, Sequeira L, Sanches M, Kemp J, et al. The influence of electronic health record use on physician burnout: cross-sectional survey. *J Med Internet Res* 2020;22(7):e19274 [FREE Full text] [doi: [10.2196/19274](https://doi.org/10.2196/19274)] [Medline: [32673234](https://pubmed.ncbi.nlm.nih.gov/32673234/)]
5. Melnick ER, Dyrbye LN, Sinsky CA, Trockel M, West CP, Nedelec L, et al. The association between perceived electronic health record usability and professional burnout among US physicians. *Mayo Clin Proc* 2020;95(3):476-487 [FREE Full text] [doi: [10.1016/j.mayocp.2019.09.024](https://doi.org/10.1016/j.mayocp.2019.09.024)] [Medline: [31735343](https://pubmed.ncbi.nlm.nih.gov/31735343/)]
6. Ontario Medical Association. Healing the healers: system-level solutions to physician burnout. Canadian Medical Association. Toronto, ON: Ontario Medical Association; 2021. URL: <https://www.cma.ca/physician-wellness-hub/resources/burnout/system-level-solutions-to-burnout> [accessed 2021-11-30]
7. Matsuo T, Kobayashi D, Taki F, Sakamoto F, Uehara Y, Mori N, et al. Prevalence of health care worker burnout during the coronavirus disease 2019 (COVID-19) pandemic in Japan. *JAMA Netw Open* 2020;3(8):e2017271 [FREE Full text] [doi: [10.1001/jamanetworkopen.2020.17271](https://doi.org/10.1001/jamanetworkopen.2020.17271)] [Medline: [32749466](https://pubmed.ncbi.nlm.nih.gov/32749466/)]
8. Jalili M, Niroomand M, Hadavand F, Zeinali K, Fotouhi A. Burnout among healthcare professionals during COVID-19 pandemic: a cross-sectional study. *Int Arch Occup Environ Health* 2021;94(6):1345-1352 [FREE Full text] [doi: [10.1007/s00420-021-01695-x](https://doi.org/10.1007/s00420-021-01695-x)] [Medline: [33864490](https://pubmed.ncbi.nlm.nih.gov/33864490/)]
9. National Academies of Sciences, Engineering, and Medicine, National Academy of Medicine, Committee on Systems Approaches to Improve Patient Care by Supporting Clinician Well-Being. Taking action against clinician burnout: a systems approach to professional well-being. Washington, D.C: National Academies Press (US); 2019.
10. Shah T, Borondy Kitts A, Gold JA, Horvath K, Ommaya A, Opelka F, et al. Electronic health record optimization and clinician well-being: a potential roadmap toward action. *NAM Perspect* 2020;1-10. [doi: [10.31478/202008a](https://doi.org/10.31478/202008a)]
11. Clickbusters. Vanderbilt Clinical Informatics Centre. URL: <https://www.vumc.org/vcl/clickbusters> [accessed 2021-11-30]
12. Schwartz JM, Lucas E, Moy AJ. 25X5 Survey: examining documentation burden reduction strategies during COVID-19. Columbia University. 2021. URL: <https://www.dbmi.columbia.edu/wp-content/uploads/2021/02/25X5-Survey-Examining-Documentation-Burden-Reduction-Strategies-during-COVID-19-1.pdf> [accessed 2021-11-30]
13. 2020 national survey of Canadian nurses: use of digital health technology in practice. Canada Health Infoway. 2020. URL: <https://www.infoway-inforoute.ca/en/component/edocman/resources/reports/benefits-evaluation/3812-2020-national-survey-of-canadian-nurses-use-of-digital-health-technology-in-practice> [accessed 2021-11-30]
14. Sequeira L, Almilaji K, Strudwick G, Jankowicz D, Tajirian T. EHR "SWAT" teams: a physician engagement initiative to improve Electronic Health Record (EHR) experiences and mitigate possible causes of EHR-related burnout. *JAMIA Open* 2021;4(2):ooab018 [FREE Full text] [doi: [10.1093/jamiaopen/ooab018](https://doi.org/10.1093/jamiaopen/ooab018)] [Medline: [33898934](https://pubmed.ncbi.nlm.nih.gov/33898934/)]
15. Wilkie T, Tajirian T, Stergiopoulos V. Advancing physician wellness, engagement and excellence in a mental health setting: a Canadian perspective. *Health Promot Int* 2022;37(1):daab061. [doi: [10.1093/heapro/daab061](https://doi.org/10.1093/heapro/daab061)] [Medline: [33942064](https://pubmed.ncbi.nlm.nih.gov/33942064/)]
16. 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* 2018;2(1):35-39 [FREE Full text] [doi: [10.1093/jamiaopen/ooy044](https://doi.org/10.1093/jamiaopen/ooy044)] [Medline: [31984343](https://pubmed.ncbi.nlm.nih.gov/31984343/)]
17. Paterson J, Fernandes J, Hunter K, Hubert S, Strudwick G. Embedding psychiatric risk flags within an electronic health record: initial findings and lessons learned. *Healthc Q* 2019;21(4):54-60. [doi: [10.12927/hcq.2019.25739](https://doi.org/10.12927/hcq.2019.25739)] [Medline: [30946656](https://pubmed.ncbi.nlm.nih.gov/30946656/)]

18. Electronic medical record adoption model (EMRAM). Healthcare Information and Management Systems Society. 2020. URL: <https://www.himss.org/what-we-do-solutions/digital-health-transformation/maturity-models/electronic-medical-record-adoption-model-emram> [accessed 2021-11-30]
19. Adoption model for analytics maturity (AMAM). Healthcare Information and Management Systems Society. 2020. URL: <https://www.himss.org/what-we-do-solutions/digital-health-transformation/maturity-models/adoption-model-analytics-maturity-amam> [accessed 2021-11-30]
20. The centre for addiction and mental health | HIMSS Davies enterprise award. Healthcare Information and Management Systems Society. 2019. URL: <https://www.himss.org/resources/centre-addiction-and-mental-health-himss-davies-enterprise-award> [accessed 2021-11-30]
21. Gawande A. Why doctors hate their computers. *The New Yorker*. 2018. URL: <https://www.newyorker.com/magazine/2018/11/12/why-doctors-hate-their-computers> [accessed 2021-11-30]
22. Collier R. Electronic health records contributing to physician burnout. *CMAJ* 2017;189(45):E1405-E1406 [FREE Full text] [doi: [10.1503/cmaj.109-5522](https://doi.org/10.1503/cmaj.109-5522)] [Medline: [29133547](https://pubmed.ncbi.nlm.nih.gov/29133547/)]
23. Kahn D, Stewart E, Duncan M, Lee E, Simon W, Lee C, et al. A prescription for note bloat: an effective progress note template. *J Hosp Med* 2018;13(6):378-382. [doi: [10.12788/jhm.2898](https://doi.org/10.12788/jhm.2898)] [Medline: [29350222](https://pubmed.ncbi.nlm.nih.gov/29350222/)]
24. Lipton M, Press R. Electronic progress notes - avoiding note bloat and other pitfalls. NYU Langone Health. 2010. URL: <https://www.bestcollections.org/wp-content/uploads/2020/04/electronic-progress-note-template.pdf> [accessed 2021-11-30]
25. Lo B, Kemp J, Cullen C, Tajirian T, Jankowicz D, Strudwick G. Electronic health record-related burnout among clinicians: practical recommendations for Canadian healthcare organizations. *Healthc Q* 2020;23(3):54-62. [doi: [10.12927/hcq.2020.26332](https://doi.org/10.12927/hcq.2020.26332)] [Medline: [33243367](https://pubmed.ncbi.nlm.nih.gov/33243367/)]
26. Kannampallil T, Abraham J, Lou SS, Payne PR. Conceptual considerations for using EHR-based activity logs to measure clinician burnout and its effects. *J Am Med Inform Assoc* 2021;28(5):1032-1037 [FREE Full text] [doi: [10.1093/jamia/ocaa305](https://doi.org/10.1093/jamia/ocaa305)] [Medline: [33355360](https://pubmed.ncbi.nlm.nih.gov/33355360/)]
27. Donnelly P, Kirk P. Use the PDSA model for effective change management. *Educ Prim Care* 2015;26(4):279-281. [doi: [10.1080/14739879.2015.11494356](https://doi.org/10.1080/14739879.2015.11494356)] [Medline: [26253069](https://pubmed.ncbi.nlm.nih.gov/26253069/)]
28. Sieja A, Markley K, Pell J, Gonzalez C, Redig B, Kneeland P, et al. Optimization sprints: improving clinician satisfaction and teamwork by rapidly reducing electronic health record burden. *Mayo Clin Proc* 2019;94(5):793-802 [FREE Full text] [doi: [10.1016/j.mayocp.2018.08.036](https://doi.org/10.1016/j.mayocp.2018.08.036)] [Medline: [30824281](https://pubmed.ncbi.nlm.nih.gov/30824281/)]
29. Sieja A, Kim E, Holmstrom H, Rotholz S, Lin CT, Gonzalez C, et al. Multidisciplinary sprint program achieved specialty-specific EHR optimization in 20 clinics. *Appl Clin Inform* 2021;12(2):329-339. [doi: [10.1055/s-0041-1728699](https://doi.org/10.1055/s-0041-1728699)] [Medline: [33882586](https://pubmed.ncbi.nlm.nih.gov/33882586/)]
30. Lo B, Almilaji K, Jankowicz D, Sequeira L, Strudwick G, Tajirian T. Application of the i-PARIHS framework in the implementation of speech recognition technology as a way of addressing documentation burden within a mental health context. In: American Medical Informatics Association 2021 Annual Symposium. 2021 Presented at: AMIA '21; October 30-November 3, 2021; San Diego, CA.
31. Berg S. Peer-led EHR training saves physicians' scarcest resource: time. American Medical Association. 2018. URL: <https://www.ama-assn.org/practice-management/digital/peer-led-ehr-training-saves-physicians-scarcest-resource-time> [accessed 2021-11-30]
32. Fernandes J, Brunton I, Strudwick G, Banik S, Strauss J. Physician experience with speech recognition software in psychiatry: usage and perspective. *BMC Res Notes* 2018;11(1):690 [FREE Full text] [doi: [10.1186/s13104-018-3790-y](https://doi.org/10.1186/s13104-018-3790-y)] [Medline: [30285818](https://pubmed.ncbi.nlm.nih.gov/30285818/)]
33. Michalik HJ, Carolan HT, Haut ER, Lau BD, Streiff MB, Finkelstein J, et al. Use of provider-level dashboards and pay-for-performance in venous thromboembolism prophylaxis. *J Hosp Med* 2015;10(3):172-178 [FREE Full text] [doi: [10.1002/jhm.2303](https://doi.org/10.1002/jhm.2303)] [Medline: [25545690](https://pubmed.ncbi.nlm.nih.gov/25545690/)]
34. Halpern DJ, Clark-Randall A, Woodall J, Anderson J, Shah K. Reducing imaging utilization in primary care through implementation of a peer comparison dashboard. *J Gen Intern Med* 2021;36(1):108-113 [FREE Full text] [doi: [10.1007/s11606-020-06164-8](https://doi.org/10.1007/s11606-020-06164-8)] [Medline: [32885372](https://pubmed.ncbi.nlm.nih.gov/32885372/)]
35. Perron CE, Bachur RG, Stack AM. Development, implementation, and use of an emergency physician performance dashboard. *Clin Pediatr Emerg Med* 2017;18(2):115-123. [doi: [10.1016/j.cpem.2017.04.004](https://doi.org/10.1016/j.cpem.2017.04.004)]
36. Yigitbasioglu OM, Velcu O. A review of dashboards in performance management: implications for design and research. *Int J Account Inf Syst* 2012;13(1):41-59. [doi: [10.1016/j.accinf.2011.08.002](https://doi.org/10.1016/j.accinf.2011.08.002)]
37. Lau F, Kuziemy C. Handbook of eHealth evaluation: an evidence-based approach. Victoria, BC: University of Victoria; 2017.
38. Leis JA, Shojania KG. A primer on PDSA: executing plan-do-study-act cycles in practice, not just in name. *BMJ Qual Saf* 2017;26(7):572-577. [doi: [10.1136/bmjqs-2016-006245](https://doi.org/10.1136/bmjqs-2016-006245)] [Medline: [27986900](https://pubmed.ncbi.nlm.nih.gov/27986900/)]
39. Lights On Network®. Cerner Corporation. 2021. URL: <https://www.cerner.com/solutions/lights-on-network> [accessed 2021-11-30]

40. Strategy on reducing regulatory and administrative burden relating to the use of health IT and EHRs. The Office of the National Coordinator for Health Information Technology. 2020. URL: <https://www.healthit.gov/topic/usability-and-provider-burden/strategy-reducing-burden-relating-use-health-it-and-ehrs> [accessed 2021-11-30]
41. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A tipping point for measurement-based care. *Psychiatr Serv* 2017;68(2):179-188. [doi: [10.1176/appi.ps.201500439](https://doi.org/10.1176/appi.ps.201500439)] [Medline: [27582237](https://pubmed.ncbi.nlm.nih.gov/27582237/)]
42. 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;76(3):324-335 [FREE Full text] [doi: [10.1001/jamapsychiatry.2018.3329](https://doi.org/10.1001/jamapsychiatry.2018.3329)] [Medline: [30566197](https://pubmed.ncbi.nlm.nih.gov/30566197/)]
43. Schwamm LH, Estrada J, Erskine A, Licurse A. Virtual care: new models of caring for our patients and workforce. *Lancet Digit Health* 2020;2(6):e282-e285 [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30104-7](https://doi.org/10.1016/S2589-7500(20)30104-7)] [Medline: [32382724](https://pubmed.ncbi.nlm.nih.gov/32382724/)]
44. Mjåset C, Lawrence K, Lee T. Hybrid physicians create 'Social Capital' for health care. *NEJM Catalyst*. 2020. URL: <https://catalyst.nejm.org/doi/full/10.1056/CAT.20.0271> [accessed 2022-02-18]

Abbreviations

CMIO: Chief Medical Information Officer

EHR: electronic health record

IT: information technology

Edited by T Leung; submitted 10.08.21; peer-reviewed by A Sieja, L Maaß, R Chan; comments to author 01.10.21; revised version received 24.11.21; accepted 07.01.22; published 08.03.22.

Please cite as:

Tajirian T, Jankowicz D, Lo B, Sequeira L, Strudwick G, Almilaji K, Stergiopoulos V

Tackling the Burden of Electronic Health Record Use Among Physicians in a Mental Health Setting: Physician Engagement Strategy
J Med Internet Res 2022;24(3):e32800

URL: <https://www.jmir.org/2022/3/e32800>

doi: [10.2196/32800](https://doi.org/10.2196/32800)

PMID: [35258473](https://pubmed.ncbi.nlm.nih.gov/35258473/)

©Tania Tajirian, Damian Jankowicz, Brian Lo, Lydia Sequeira, Gillian Strudwick, Khaled Almilaji, Vicky Stergiopoulos. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 08.03.2022. 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.

Viewpoint

A National Network of Safe Havens: Scottish Perspective

Chuang Gao¹, PhD; Mark McGilchrist¹, PhD; Shahzad Mumtaz¹, PhD; Christopher Hall¹, BSc; Lesley Ann Anderson², PhD; John Zurowski³, BSc; Sharon Gordon⁴, PhD; Joanne Lumsden⁴, PhD; Vicky Munro⁴, BSc; Artur Wozniak⁴, BSc; Michael Sibley⁵, MSc; Christopher Banks⁵, BSc; Chris Duncan⁶, BSc; Pamela Linksted⁶, BSc; Alastair Hume⁷, BSc; Catherine L Stables⁸, PhD; Charlie Mayor⁹, PhD; Jacqueline Caldwell⁵, MBA; Katie Wilde⁴, MMath; Christian Cole¹, PhD; Emily Jefferson¹, BSc, PhD

¹Health Informatics Centre, Ninewells Hospital & Medical School, University of Dundee, Dundee, United Kingdom

²Centre for Health Data Science, University of Aberdeen, Aberdeen, United Kingdom

³Imaging Centre of Excellence, Queen Elizabeth University Hospital, Glasgow, United Kingdom

⁴Grampian Data Safe Haven, Aberdeen Centre for Health Data Science, University of Aberdeen, Aberdeen, United Kingdom

⁵Electronic Data Research and Innovation Service, Public Health Scotland, Edinburgh, United Kingdom

⁶Lothian Research Safe Haven, Department of Public Health and Health Policy National Health Service Lothian, Edinburgh, United Kingdom

⁷EPCC, University of Edinburgh, Edinburgh, United Kingdom

⁸DataLoch, Usher Institute, University of Edinburgh, Edinburgh, United Kingdom

⁹Glasgow Safe Haven, Research and Development division of National Health Service Greater Glasgow and Clyde, Glasgow, United Kingdom

Corresponding Author:

Emily Jefferson, BSc, PhD

Health Informatics Centre

Ninewells Hospital & Medical School

University of Dundee

Mail Box 15

Dundee, DD1 9SY

United Kingdom

Phone: 44 (0)1382 383943

Email: e.r.jefferson@dundee.ac.uk

Abstract

For over a decade, Scotland has implemented and operationalized a system of Safe Havens, which provides secure analytics platforms for researchers to access linked, deidentified electronic health records (EHRs) while managing the risk of unauthorized reidentification. In this paper, a perspective is provided on the state-of-the-art Scottish Safe Haven network, including its evolution, to define the key activities required to scale the Scottish Safe Haven network's capability to facilitate research and health care improvement initiatives. A set of processes related to EHR data and their delivery in Scotland have been discussed. An interview with each Safe Haven was conducted to understand their services in detail, as well as their commonalities. The results show how Safe Havens in Scotland have protected privacy while facilitating the reuse of the EHR data. This study provides a common definition of a Safe Haven and promotes a consistent understanding among the Scottish Safe Haven network and the clinical and academic research community. We conclude by identifying areas where efficiencies across the network can be made to meet the needs of population-level studies at scale.

(*J Med Internet Res* 2022;24(3):e31684) doi:[10.2196/31684](https://doi.org/10.2196/31684)

KEYWORDS

electronic health records; Safe Haven; data governance

Introduction

Background

Electronic health records (EHRs) are routinely collected data that are generated when an individual receives care in a health

care setting. EHRs typically contain records of medical history, diagnoses, medications, allergies, immunizations, other treatments, and laboratory results [1]. The records can be generated in different settings (eg, primary care facilities, such as clinics and health care centers, and secondary care facilities, such as hospitals and emergency care centers). Although the

primary purpose of EHRs is to improve the direct care of patients, they also have some other purposes that are termed *secondary use* or *reuse* [2]. Using EHR data in research is one such type of secondary use [3,4].

Safe Havens are secure environments that have been widely used to support access to EHRs for research while protecting patient identity and privacy [5,6]. The 4 Safe Havens collaborating as part of the UK-wide Farr Institute were described by Lea et al [5] and were found to have different processes, controls, and environments. In Scotland, a network of 5 Safe Havens has been established to support EHR reuse and, over the past decade, has enabled researchers to access data at scale [6].

The Scottish network of Safe Havens has been highly successful in supporting research. Over the past 5 years, the network has supported >1000 research studies. There are a small number of research and innovation projects (eg, the Industrial Center for Artificial Intelligence Research in Digital diagnostics [7] and Research Data Scotland [8]) that are collaborations across Safe Havens. However, most research projects are delivered by a single Safe Haven. Each Safe Haven maintains and controls access to EHR data collected from their geographically local regions and therefore has detailed knowledge of these data sets. The exception in Scotland is the national Safe Haven (electronic Data Research and Innovation Service [eDRIS]), which holds national-level data sets. Researchers generally only access either the breadth of the nationally held data, with high cohort coverage, which are collected at a Scottish level, or the depth of the local clinical data, which has more detailed information about persons or entities from the regional Safe Havens.

Representatives from each Safe Haven within the network meet regularly and are supported and chaired by the Scottish government's Chief Scientist Office. The Safe Havens collaborate to develop and share best practices. The network is primarily funded on a cost-recovery basis by charging researchers for services, with some Safe Havens also receiving

some core support from the National Health Service (NHS) Scotland Research and Development funds.

This study provides an analysis of the infrastructure, operations, and features of each Safe Haven and assesses how these affect the interoperability and technical options to support multi-Safe Haven projects. We present how Safe Havens in Scotland have protected privacy and facilitated the reuse of the EHR data.

What Is a Safe Haven?

Overview

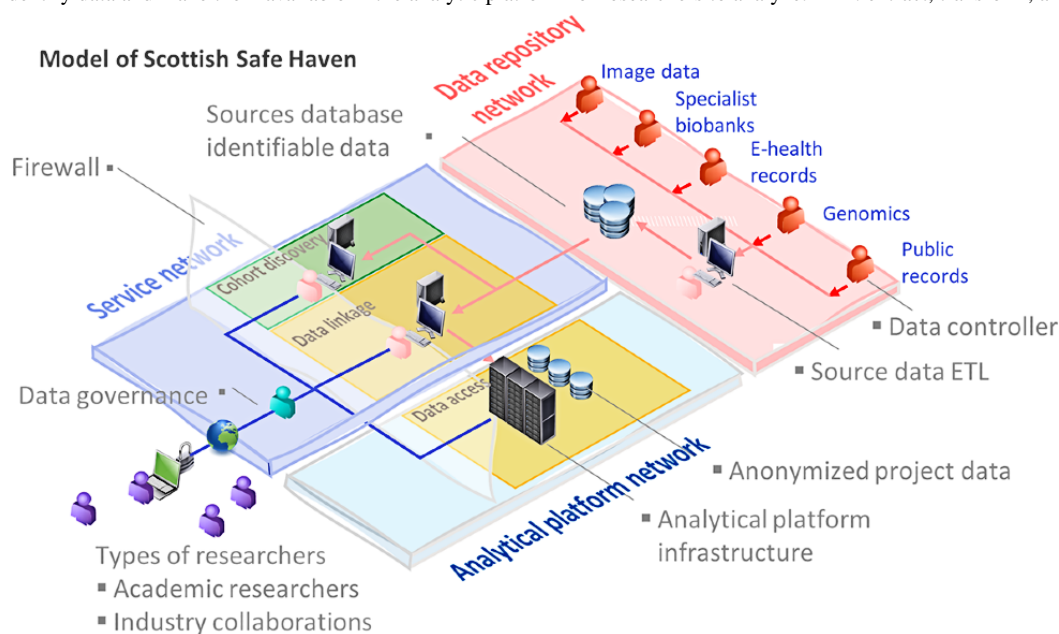
Safe Havens have evolved as a set of processes for supporting researchers accessing sensitive data in a streamlined and secure way while maintaining patient confidentiality [5,9,10]. The term *SafeHaven* is widely used but can have different meanings to different people and in different contexts. Barton et al [11] described, in detail, the origins and evolution of the term. A Safe Haven was defined as follows:

A repository in which useful but potentially sensitive data may be kept securely under governance and informatics systems that are fit-for-purpose and appropriately tailored to the nature of the data being maintained, and may be accessed and utilised by legitimate users undertaking work and research contributing to biomedicine, health and/or to the ongoing development of healthcare systems.

Concerning health data, other Safe Havens or similar infrastructures [5,11] exist nationally [12-14] and internationally [15-17]. Kavianpour et al [18] provided a review of trusted research environments based on the interviews of 20 UK national and international Safe Havens. This paper provides a perspective on Safe Havens in Scotland and is based upon the direct experiences of the authors.

Figure 1 provides a model of how Scottish Safe Havens are structured. We have identified that Scottish Safe Havens mainly offer services that are described in the following sections.

Figure 1. Model of Scottish Safe Havens. Researchers have access to the Safe Haven application process after data governance approvals. Safe Haven staff link and deidentify data and make them available in the analytic platform for researchers to analyze. ETL: extract, transform, and load.



A Data Processor and Data Repository Management

This involves the secure handling and linking of data from multiple sources and possible hosting or managing of longitudinal data (detailed information about persons or entities, such as conditions, hospital admissions, and prescription data). Scottish Safe Havens can also provide the function of a *trusted third party* [6,19]. They can support the linkage of identifiable information where the roles of *indexer* and *linker* (see detailed definition in *Data Linkage* section) are separated so that no single organization or individual has visibility of another organization's identifiable data linked to their descriptive data [20,21]. Safe Havens function as data processors [6] for any given data set and agree on terms with each data controller (Safe Havens can also be the data controller) to ensure that activities are centrally logged, monitored, and audited [6].

Analytical Platform

An analytical platform is a highly secure, high-performance computing environment that enables researchers to securely analyze data without the row-level deidentified data leaving the environment (only aggregate level results can be exported). Strict governance and controls are implemented to ensure data security in the analytical platform.

Research Support

The Safe Haven coordinators provide support to researchers navigating the data requirements and permissions landscape and provide a review mechanism to share the lessons from one project to the next. Some Scottish Safe Havens provide support for analysis. Internal Safe Haven data scientists can help the research group with statistical analysis.

The term *Safe Haven* is defined here as the overarching service that combines the previously mentioned services: a data processor and data repository, an analytics platform, and research support.

The Scottish Safe Havens follow the *five safes* principles of a trusted research environment—safe people, safe project, safe setting, safe data, and safe output [14]—as described within the Health Data Research United Kingdom green paper [14].

Scottish Federated Network of Safe Havens

The network of 5 Safe Havens operating in Scotland is accredited by the Scottish government and each Safe Haven adheres to the Scottish Safe Haven Charter [6]. Each offers the 3 services, which are described in the *What is a Safe Haven?* section, with different data access procedures (subject to the necessary local governance approvals), applied to different data sources and with different standard operating procedures.

There are 4 regional Safe Havens and 1 national Safe Haven. There is a regional Safe Haven for each research and development node of the NHS supported by the Scottish government's Chief Scientist Office [22]. They are provisioned by partnerships between the NHS boards within each research and development node and with a leading university from the region. Whether the primary contact organization for a Safe Haven is an NHS board or a university differs between regional Safe Havens (Table 1). eDRIS [23], part of Public Health Scotland (PHS) [24], commissions the Edinburgh Parallel Computing Centre (EPCC) [25], University of Edinburgh, to provide the national Safe Haven. Grampian Data Safe Haven (DaSH) [26], a collaboration between the University of Aberdeen and NHS Grampian, is the Safe Haven for the Grampian region encompassing Aberdeen City, Aberdeenshire, and Moray. The Health Informatics Center (HIC) [27] at the University of Dundee covers the Tayside and Fife regions. The Glasgow [28] and Lothian or DataLoch Safe Havens [29,30] are led by the NHS, covering the west of Scotland, Edinburgh, and the South East region, and working in collaboration with the Glasgow and Edinburgh Universities, respectively.

Table 1. A summary table of Safe Haven properties.

Function	Safe Haven				
	eDRIS ^a (national)	DaSH ^b	Glasgow Safe Haven	HIC ^c	Lothian or DataLoch ^d
General and data governance					
Safe Haven affiliation	PHS ^e	UoA ^f or NHS ^g	NHS	UoD ^h or NHS	NHS
Analytical platform affiliation	UoE ⁱ (EPCC ^j)	UoA	UoG ^k (RCB ^l)	UoD	UoE (EPCC)
Network for Safe Haven services (cohort building and linkage)	NHSnet or EPCC	NHSnet	NHSG ^m or NHSnet	NHS	NHSL ⁿ or NHSnet
Network for analytical platform	UoE or Janet	UoA or Janet	UoG or Janet	UoD or Janet and secure public cloud	UoE or Janet
Data repository network	NHSnet or EPCC	NHSnet	NHSnet	NHSnet	NHSnet
Geographical region ^o	Scotland	NHS Grampian	West of Scotland	NHS Tayside and Fife	Lothian or South East of Scotland
Population ^p	5.7 million	600,000	1.2 million	850,000	900,000
Active projects in 2020	>600	>120	>100	>100	>20
Controller or controllers	PHS+NRS ^q +Scottish government	Original data sources	Original data sources	Original data sources	Original data sources
Processor or processors	eDRIS	DaSH	Glasgow Safe Haven	HIC	Lothian or DataLoch
Governance committee	Health and Social Care PBPP ^r and Statistics PBPP	North Node Privacy Advisory Committee	Privacy advisory committee	HIC governance committee	Data access committee
Data discovery or metadata					
Feasibility	Manual or NDC ^s	Manual or local documents	Manual, local documents, or TriNetX ^t	Manual or using RDMP ^u [19] automation	Manual or local documents
Metadata provided with project extracts	No	Yes or standard (workflow)	Yes or bespoke	Yes or standard (RDMP)	Yes or bespoke
Phenotype or cohort development	ICD ^v code from user	By user	Locally stored algorithms or user	By user	By user or CALIBER Library
Data linkage and dedeidentification					
Indexer	External (PHS for health data)	Internal	Internal	Internal (RDMP)	Internal
Deidentification method	Workflow [31]	SQL procedure	Database views (usually SQL)	Workflow (RDMP)	SQL procedure
CHI ^w seeding	NSS ^x or CHI Linkage Team	CHI Linkage Team or internal	Internal	Internal	CHI Linkage Team
Analytic platform					
Archival	NHSnet and UoE (EPCC)	UoA	NHSnet and UoG (RCB)	NHSnet and UoD and secure public cloud	NHSnet and UoE (EPCC)
Project data content standards	As source	As source or ICD	As source	As source	As source
Project data format standards	CSV	SPSS, Stata, or CSV	CSV	CSV or database	CSV
Data repository					

Function	Safe Haven				Lothian or DataLoch ^d
	eDRIS ^a (national)	DaSH ^b	Glasgow Safe Haven	HIC ^c	
Data repository number	≥85	1	1	1	1 each
Data repository ownership	No	Yes	Yes	Yes	Yes
Source data metadata	NDC	Internal shared files	Internal shared files	RDMP	Data dictionaries
Metadata publicly available	Yes	No	No	Yes	Yes ^y
Number of data sets available	85	40	≥200	163	12
Source data extract, transform, and load	Data management team PHS	Internal (SQL and Python)	Business Intelligence and Informatics in NHSG	Internal (RDMP)	Internal (SQL and Python)
Repository uses CDM ^z	No (proprietary)	No (proprietary)	No (proprietary)	No (proprietary)	No (proprietary)

^aeDRIS: electronic Data Research and Innovation Service.

^bDaSH: Grampian Data Safe Haven.

^cHIC: Health Informatics Centre.

^dWhen this work was conducted, the Lothian Research Safe Haven (LRSH) and DataLoch were separate (though closely partnered). Since April 1, 2021, LRSH has been integrated within the DataLoch service.

^ePHS: Public Health Scotland.

^fUoA: University of Aberdeen.

^gNHS: National Health Service.

^hUoD: University of Dundee.

ⁱUoE: University of Edinburgh.

^jEPCC: Edinburgh Parallel Computing Centre.

^kUoG: University of Glasgow.

^lRCB: Robertson Centre For Biostatistics.

^mNHSG: National Health Service Glasgow.

ⁿNHSL: National Health Service Lothian.

^oRegional Safe Havens have governance to request regional health board data. For example, Glasgow Safe Haven can request West of Scotland Health Board data.

^pSafe Havens have access to historic records for patients who are deceased, which can increase the accessible data.

^qNRS: National Records Scotland.

^rPBPP: Public Benefit And Privacy Panel.

^sNDC: national data catalog.

^tTriNetX is a health research network tool that connects to assist drug discovery by helping pharmaceutical companies access clinical data. Glasgow Safe Haven has a TriNetX node. For data mapped into TriNetX tool, their study feasibility can be done using TriNetX.

^uRDMP: Research Data Management Platform.

^vICD: International Statistical Classification of Diseases and Related Health Problems.

^wCHI: community health index.

^xNSS: National Services Scotland.

^yCOVID-19 data dictionary is on DataLoch website.

^zCDM: common data model.

Scottish NHS Data Sources

Overview

Scotland has a single health care provider (NHS Scotland) and world-leading national health-linked data assets from birth to death. In a high-level summary, the national Safe Haven has direct access to health administrative data, with high cohort coverage collected at a Scottish national level, and the regional Safe Havens have direct access to more detailed health data

from clinical systems. Regional Safe Havens can work closely with local data custodians, which gives them easy access to additional data sources that are not routinely held (eg, other health data, educational data, or police data). Access to these other sources of data may require additional time because of different access processes and governance approvals.

The Research Data Scotland initiative [32] has been set up to streamline and support access to linked health and administrative data sets across the country.

National-Level NHS Data

The PHS collects national-level NHS [33] and administrative data to provide health information, health intelligence, statistical services, and advice to support the NHS in progressing quality improvement in health and care and facilitate robust planning and decision-making. These data sets can be accessed in the national Safe Haven. Each health board across Scotland provides a regular update of a subset of their identifiable administrative data to PHS. This is standardized by PHS to create homogeneous data within several national databases. Such data include Scottish Morbidity Records (SMRs) and community-dispensed prescriptions. SMR data cover several different data sets such as SMR00 (hospital outpatient), SMR01 (acute stay hospital admissions), SMR02 (maternity), SMR04 (psychiatric returns), SMR06 (cancer registry), SMR11 (neonatal), and SMR25 (substance misuse). National Records Scotland (NRS) records births, marriages, and deaths.

Prescription data are collected nationally in 2 different ways. Through the e-Pharmacy system [34], prescriptions written by general practitioners are captured directly in the system. The long-standing Data Capture Validation and Pricing paid system [35] is used in Scotland to capture dispensing data that determine remuneration for community pharmacies. The *watermarked* prescriptions go from general practitioners to the patient and then to a pharmacy and are then collected and transferred monthly to Data Capture Validation and Pricing or PHS for automated processing [36].

Regional-Level NHS Data

The regional Safe Havens all receive a subset of the data from the national standardized data sets (eg, SMR and prescribing data) from PHS, which includes only the patients who are residents or received health care within the relevant boards. They also have access to the deeply phenotyped data that are captured within local clinical systems but not collected at a national level. For example, they have access to the following data: microbiology, virology, laboratory test, stroke, and echocardiology. The type and level of available local data differ between Safe Havens. Individuals in Scotland are assigned a community health index (CHI) number [37] when they first interact with the health service. This is retained within their EHRs as much as possible throughout their health history. Regional Safe Havens use CHIs to link data sets to the nationally captured records for the population within their region.

Research Data

In addition to unconsented access to routinely collected administrative or clinical records, the national and regional Safe Havens can also host or manage researcher-collected consented data sets from many sources such as clinical trials and patient questionnaires. Compared with routinely collected EHRs, the research data often cover a narrower spectrum but provide more detailed information about the individual. Participants in research cohorts are volunteers who have consented to data access rules approved by ethics at the outset of the study. For example, Generation Scotland [38] is a resource of human biological samples and data that are available for medical research to create more effective treatments based on gene

knowledge for the health, social, and economic benefit of Scotland and its people. Another example is the Scottish Health Research Register [39] cohort, in which >280,000 individuals consented and were recruited to allow for the genotyping of any remaining blood samples after routine tests and applied them to research on their health data [39].

Regional Safe Havens also host disease-specific study data [40-45]. The data within these studies can be collected from a range of sources: clinical data, patient surveys, and routinely collected EHR data.

Some disease registries were originally created at a regional level but were then rolled out nationally. For example, the Scottish Care Information-Diabetes (SCI-Diabetes) [46] disease registry was formed by curating and linking routinely collected data from the Tayside Region. It was later developed into a nationwide resource that now collects patient-reported outcome data. The data collected in SCI-Diabetes are used in clinical care. Extracts from SCI-Diabetes can also be linked on a study-by-study basis for research studies by the regional or the national Safe Havens.

Safe Havens can link research data to routinely collected administrative or clinical records and provide access to the combined data (in a deidentified form) within an analytical platform for analysis.

Scottish Safe Havens

Overview

Each Scottish Safe Haven has its data repository hosted on the NHS network (Table 1), except for the national Safe Haven, which also hosts some data within the EPCC on a secure university environment. Safe Havens have data-sharing agreements with multiple data controllers and regularly receive new data from them.

All Safe Havens have committed to an approach to data access based on analytical platforms. Each Safe Haven has either established or has access to an analytical platform.

There are differences among Safe Havens in how they achieve the 3 main services described. Table 1 summarizes each Scottish Safe Haven. In the following section, we discuss the Safe Havens in detail and their common deployment features.

Data Governance and Workflow

The governance approval step looks into aspects of the project such as ethics, peer review, funding source, public benefit, and adherence to the *five safes*, as described in the *What is a Safe Haven?* section. The governance approval process varies among Safe Havens. Even for the same Safe Haven, different projects may require different governance approval processes to satisfy the different data controllers. However, the governance process for a standard deidentified project where data are accessed in an analytical platform is relatively streamlined. Each Safe Haven has a delegated governance authority or committee, as shown in Table 1, which is relatively fast and includes representatives from sponsors, ethics, lay members, and the NHS board to streamline the governance process. For the DaSH Safe Haven, projects with local researchers using local data can obtain

governance approval through the North Node Privacy Advisory Committee process. The HIC Safe Haven's *standard projects* (where deidentified data are analyzed by approved academic researchers within the analytical platform and the activity is funded by a peer-reviewed research grant) are covered by a blanket governance approval. The list of the supported standard projects is provided to the relevant governance committee for information rather than a request for approval for each one before the research commences. For Lothian or DataLoch, projects involving deidentified data within an analytical platform can obtain governance approval through their local data access committee process, which includes delegated Caldicott review.

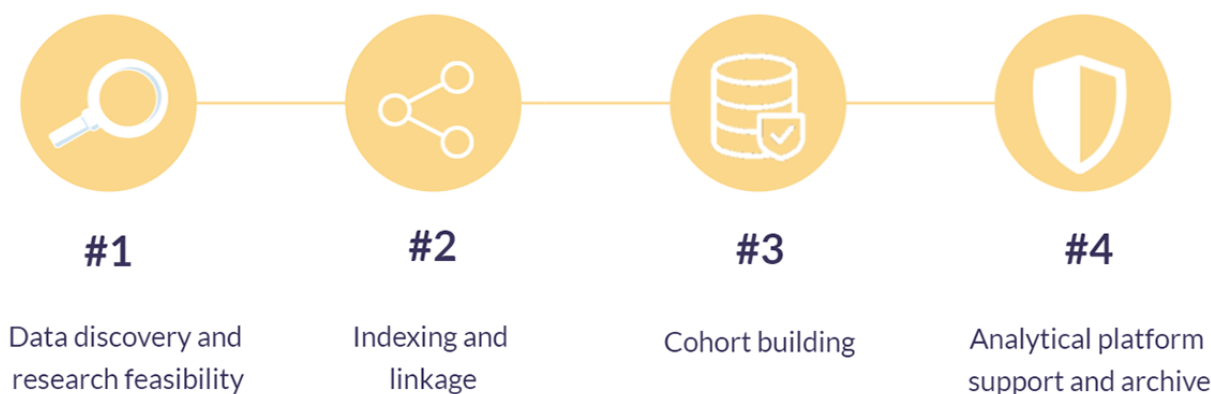
Most Safe Haven responses to project requests adhere to a standard set of processes (eg, deidentified linked data are provided within an analytical platform for academic research). In exceptional circumstances, some projects require a different model, and such exceptions need to be justified to obtain governance approvals. Example exceptions include the prepared project data not being placed in an analytical platform and the project data having some identifiable information.

To work on identifiable EHR data within a data repository, Safe Haven staff members are either NHS employees or have honorary NHS contracts. All Safe Havens have rule-based segregation of the teams, specifying those with and without

access to identifiable data. Only a handful of people in each Safe Haven can access the NHS network and see identifiable data. Other data sources (eg, administrative data generated by the government or research data generated by research institutions) can be linked to EHRs. The linkage is performed by the Safe Haven data linkage team, and the linked data sets are then hosted on the analytical platform for the approved researchers or investigators to access. At each stage, there is an oversight step to ensure all procedures are correctly followed and no unintentional identifiable data are released.

The project workflow for a data request is consistent across the Safe Haven network, as shown in [Figure 2](#). In the first step, the Safe Haven team runs research feasibility queries to identify the data needed for the research topic. Once funding and governance are in place, data linkage is conducted (as required). Data extracted from the NHS network are deidentified, validated, and assessed for disclosure before being released into the analytical platform. The details of linking and deidentification are given in the *Data Linkage* and *Data Deidentification* sections. The section *Data Formats in the Analytical Platform* provides discussions on the analytical platform support and data support. The archiving procedure of each Safe Haven and the infrastructures of each Safe Haven's data repository are discussed in the *Data Repository Infrastructure* section.

Figure 2. The Safe Haven project workflow describes the stages a Safe Haven takes to support a typical project. (1) Data discovery and research feasibility—users will initialize the application on the data governance aspects; (2) (optionally) index and link a research data set or administrative or clinical data set for hosting at a given analytic platform; (3) cohort building the selected or agreed data from Safe Haven data sets; (4) the transfer of extracted data to an analytic platform after the data governance has been checked; a user analyzes analytic platform data set. The project data set is archived at the end of the project.



Data Discovery and Metadata

Research feasibility analysis and data discovery remain a manual process involving discussions between researchers and the Safe Haven teams. During the project planning stage, researchers contact the relevant Safe Haven by email or phone call. Data discovery and research feasibility are conducted by document exchange or a face-to-face meeting. Safe Havens in Scotland require a meeting to capture the requirements of each study and guide the governance process. Research feasibility is conducted by the Safe Haven by generating aggregate numbers for cohort or subcohort sizes based on the requirements defined by the researcher or researchers (eg, the number of people in the data with diabetes who are aged >65 years and who regularly have a prescription for insulin).

Researchers normally specify a phenotype or public phenotype algorithms to identify the correct cohort for their study. As shown in [Table 1](#), no common standard procedure exists among the 5 Safe Havens to capture and reuse phenotype algorithms. However, DataLoch also uses the CALIBER phenotype library [47], whereas the Glasgow Safe Haven uses a suite of local matrix file storage phenotype algorithms, based on standard or published methods, which have been quality checked by clinicians. As the eDRIS mainly works on national data sets using the International Statistical Classification of Diseases and Related Health Problems (ICD) standard [48], it usually agrees on what the ICD codes are with the researcher and conducts the cohort building using the codes and any date or other constraints given by the researcher. The remaining Safe Havens—HIC and DaSH—normally rely on the researcher to define the cohort themselves, where researchers have the choice of phenotype

definition (eg, CALIBER phenotypes or ICD codes). Cohort identification is sometimes an iterative process between researchers and the Safe Haven team where a data constraint is applied, the impact on cohort size is observed, and the constraint is adjusted to optimize the cohort.

At the national level, PHS produces a national data catalog [49] as a single definitive resource of information on Scottish health and social care data sets to assist cohort discovery.

Metadata provides the semantics associated with the Safe Haven data sets. There is limited visibility of the metadata and data provenance available from regional Safe Havens. Most of the Safe Havens list the names of their easily accessible databases on the web [24,26,28-30,49,50] and provide researchers with a brief overview of the most commonly accessed data sets. Both HIC and eDRIS add their metadata and data sets to the Health Data Research Innovation Gateway [51]. There is no common structure in EHR data storage across the health care system in Scotland. As only a limited number of data scientists and analysts have experience in handling NHS data, this lack of visibility of metadata and data provenance can lead to a gap in understanding by data scientists and analysts about what data are available. Some Safe Havens will only release detailed metadata once they have an initial understanding of the project's needs. There are multiple initiatives, both internal [7,52] and external [8,51], that aim to improve the metadata visibility within the Scottish network of Safe Havens.

Research projects benefit from having clinical investigators who are familiar with NHS data or data scientists and analysts who have previous experience in working on Safe Haven projects within their project team. Such individuals can help identify what data are available and advise and support the data scientists and analysts working on the project. Most Safe Haven projects generally require a suitable sponsor with relevant expertise to take responsibility for the initiation and management of the project and support the project as an ethical safeguard.

All 5 Safe Havens provide research projects with metadata at the field level once a project is funded and approved and data extracts are provided for analysis. However, feasibility discussions will generally take place at the *conceptual* level. For example, a cohort definition may involve a fasting glucose constraint. The Safe Haven team will confirm that such a constraint is possible without disclosing the precise fields. To avoid bias and to get researchers to articulate what they need and what is available, this *conceptual*-level feasibility can be quite limiting. The same could be true of the data extract requested (eg, delivering BMI rather than height and weight). However, data extraction and delivery are generally at the field level, and field-level metadata are provided to ensure researchers can perform their analyses.

Table 1 shows that eDRIS provides metadata details on its website metadata (similar to the Cribsheet [53] on SMR), which can be used by researchers to define the fields they need when applying for data access. Researchers using regional Safe Havens can also use this metadata information for the nationally standardized data sets that the regional Safe Havens hold for the subset of their region (eg, SMR data). None of the Safe

Havens provide non-field-based metadata, such as through an ontology.

The eDRIS Safe Haven does not provide bespoke metadata to the user when delivering the project data. DaSH and HIC Safe Havens have a standard workflow and delivery format for supplying project-specific metadata to each project. In the Glasgow and Lothian or Dataloch Safe Havens, projects are provided with all available metadata and provenance information. No standard format is used, and thus, the included information varies from project to project.

Safe Havens have different approaches to storing metadata about data sets in their data repositories. For eDRIS, PHS updates and maintains the national data catalog that contains all the metadata for national data sets. The HIC Safe Haven uses an in-house, open-source software tool called the Research Data Management Platform (RDMP) [19] for importing data to their servers. The RDMP generates consistently formatted metadata for imported data sets. The Lothian or DataLoch Safe Haven provides *data dictionaries*, which include metadata, for all the data sets in their Relational Database Management System (RDBMS). The Glasgow Safe Haven and the DaSH Safe Haven have internal document spaces to host the metadata and provenance provided by various data sources, which are manually entered and updated by staff. The lack of standard procedures in the Scottish Safe Haven network has resulted in the available metadata varying among data sets. Highly processed data sets—which have gone through extract, transform, and load (ETL) procedures—have field parameters and rules imposed on them. These data sets have rich metadata associated with them. However, most clinical data are inherently of variable quality, with poor coverage and inconsistent and missing fields. The data set metadata do not typically inform the user of the data variability or quality issues in the original data.

Data Linkage

Overview

To answer many research questions, data linkage is required to enrich information about a defined cohort. Some Safe Haven projects involve linking NHS data with non-NHS data. Figure 3 illustrates the indexing and linking services in the Scotland Safe Haven network.

According to the guiding principles of data linkage [21], an indexer is defined as follows:

Individual (or body) who receives personal data from one or more Data Controllers and determines which records in each dataset relate to the same individual (or entity). The indexer creates a unique reference for each individual (or entity) and a corresponding key to allow the data from the different sources to be joined.

Thus, an indexing service [21] returns a unique identifier for each individual, given an input data set of identifying information (eg, name, address, date of birth, and other operational identifiers such as CHI number). The relationship between the identifiers associated with multiple data sets is maintained by the indexing service. The indexing service does

not have visibility of the descriptive data pertaining to any individual (also termed payload data, eg, an individual’s hospital admission information).

A linker or linking service is defined as follows: “Individual (or body) who receives datasets from data controllers and links them together using a key created by the indexer” [21].

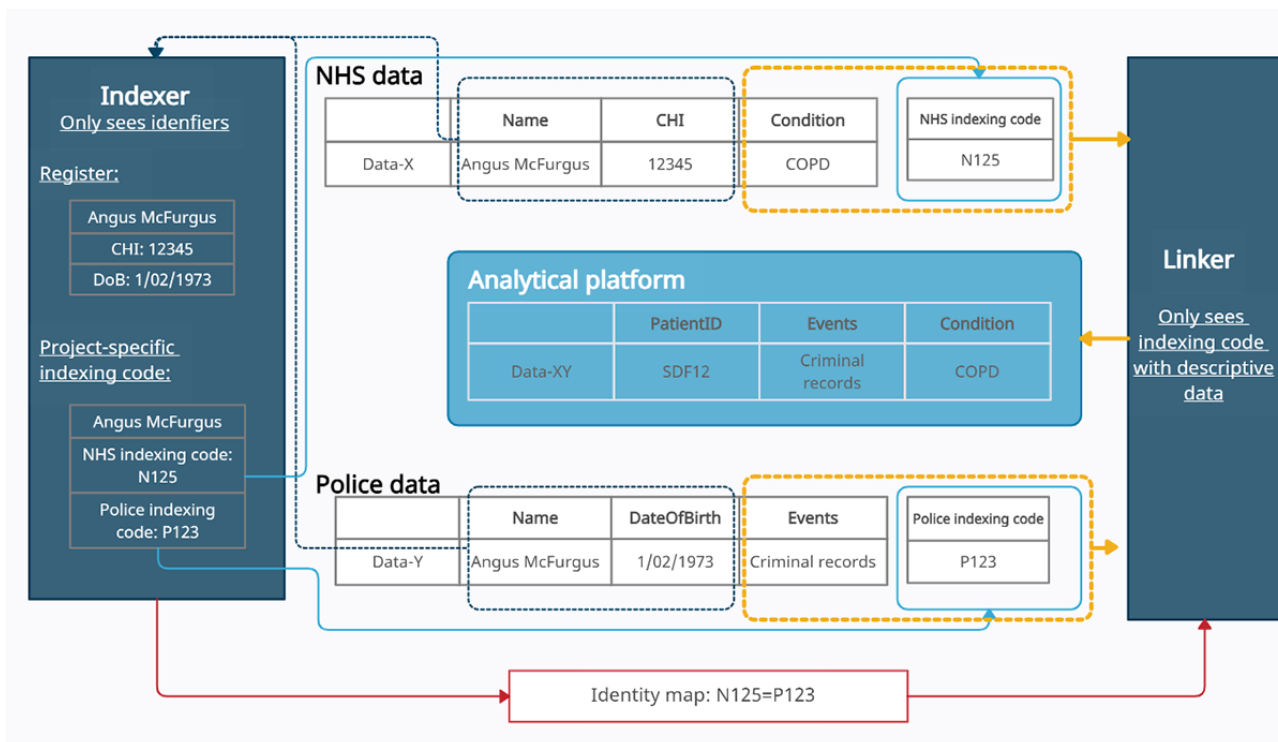
In this way, only output identifiers from the indexer service are exposed to the linker; only the linking service and the researchers see the linked data [20].

In Scotland, the NHS maintains the CHI. This is a patient identifier that concatenates a unique number, the person’s date of birth, and their sex. CHI numbers are allocated at birth or on

the first contact with the NHS in Scotland [37]. Linking health data to other data where both data sets already contain CHI numbers is straightforward. All 4 regional Safe Havens do this when preparing data for placement in an analytical platform using either software tools (eg, HIC use RDMP [19]) or RDBMS user interfaces.

The national Safe Haven, eDRIS, has established a data indexing and linkage procedure [31]. The input identifiers are personal identifiers, and the output identifier is an anonymized ID. eDRIS only receives data from providers with anonymized IDs and acts as a linker, placing the integrated data into a secure environment.

Figure 3. Data indexing and linking services in Scotland. CHI: community health index; COPD: chronic obstructive pulmonary disease; NHS: National Health Service.



CHI Seeding

When linking to a source data set that does not have CHI numbers but features other identifiers, the indexing team will use *probabilistic matching* against the population spine. Related to the CHI, the population spine [31] contains the personal identifiers of all individuals in Scotland who have been in contact with NHS Scotland. The process of matching source data sets to the population spine is known as *CHI seeding*. The recent seeding of regional social care systems with CHI is an example of this. CHI seeding is also important for historical data analysis of EHRs before the introduction of CHI indexing. In Scotland, 2 teams provide national-level CHI seeding using probability matching: the NRS Indexing Team and the PHS CHI Linkage Team (CHILI). When a research project only needs NHS data, indexing would be conducted by CHILI.

Both eDRIS and the Lothian Safe Havens rely on the NRS or CHILI for CHI seeding. DaSH Safe Haven provides CHI seeding through the NRS Scotland’s indexing team; they will

only do it themselves when they have specific personal identifiers available, such as patient name, and the data set comprises only a small amount of local Grampian patient records (approximately 500 people). Glasgow and HIC Safe Havens have a more established probabilistic matching routine developed and normally perform CHI seeding themselves. HIC has worked with local authorities to CHI seed their nonhealth data to be able to link it to health data.

Data Deidentification

Deidentification is undertaken before a Safe Haven provides the data to an analytical platform for the researcher to access. Deidentification replaces information that could identify an individual in a data set with a study identifier (ID) for that individual, which is specific to that study, or dilutes the identifier to remove its individual nature. In linked, deidentified data sets, the study ID is the same across data sources, enabling researchers to link these data sources and understand which data correspond to the same individual within that study but without

knowing their identity. This also means that deidentified IDs are unique to that project, and therefore, the same individual will have different IDs in different projects.

In general, Safe Havens apply consistent rules to identifiable data fields. Customizing deidentification rules based on the bespoke project requirements, governance approvals, and the variety of data sets can be accommodated. The treatment of identifiers depends on the project's specific justification following data minimization principles [21]. For example, a date of birth can be processed to the first day of the month or be replaced with *age-at*, or it can be removed if it is not considered necessary for the analysis. A postcode can be replaced with a deprivation score or a Scottish Index of Multiple Deprivation rank [54], or it can be removed from the data. For biometric data, where, for example, the weight and height of the individual are included, Safe Havens often put such values into ranges. Each Safe Haven follows standard operating procedures for reproducibility, consistency, and error reduction. The Scottish Safe Havens are data controllers under the European Union provisions of the General Data Protection Regulation and are individually responsible for their local data. The Scottish Safe Havens are accredited by the Scottish Government and International Organization for Standardization 27001 [55,56] on the common information security standard. The Scottish Safe Haven network has not adopted a cross-network risk of reidentification metrics [57-59]. The *five safes* principle [14]—safe people, safe project, safe setting, safe data, and safe output—ensures that the risk of reidentification is very low.

All Safe Havens indicated that they find it challenging to deidentify clinical reports and other documents containing free text, which often contain personal identifiers such as phone numbers and names. Safe Havens often exclude entire fields from research extracts when they are not confident that such fields are safe to release. The Industrial Center for Artificial Intelligence Research in Digital diagnostics [7] uses *hidden in plain sight* techniques for identifiable data on images. eDRIS has developed algorithms to remove personal identifying information from the dose instructions on the Prescribing Information System (these can also extract structured information such as dose unit and frequency). As part of the Scottish Medical Imaging service [60] and PICTURES [61], there is work in progress to deidentify and create metadata from the text written by radiologists on their findings. This uses natural language processing and the CogStack framework [62].

Data Formats in the Analytic Platform

In general, Safe Havens make few changes to the source data provided to researchers, these changes being limited to the process of deidentification. For example, there has been no attempt to harmonize data through the transformation of diagnosis codes or drug codes, where significant versioning occurs in longitudinal data. However, some Safe Havens do add derived data to data sets. Within HIC, for example, these data derivations and transformations can be applied either within the Safe Haven or at the point of extracting a deidentified research data set. This is done using RDMP, an open-source

solution that allows custom coding, or a researcher-created statistics package code to be executed in a repeatable and reproducible manner. When requested by the researcher, DaSH Safe Haven can provide the Charlson Comorbidity Index [63] and Tonelli codes [64] alongside the ICD codes. Although data standards are not applied at data extraction and delivered to an analytical platform, standards are enforced for nationally captured data sets. A team in PHS works with the health boards and system suppliers to ensure the use of standards (eg, SMR data must be structured in an agreed-upon way and use agreed-upon coding systems for content).

Safe Havens make their best efforts to accommodate the requirements of projects. However, the software available in most analytical platforms is limited (Microsoft Office packages, SPSS (IBM Corp) [65], Stata (StataCorp) [66], SAS (SAS Institute, Inc) [67], and R (R Foundation for Statistical Computing) [68]), and thus, the output data formats are also limited to R, Excel, SPSS, or Stata files. The exception is the recently launched HIC hybrid, cloud-based, scalable analytical platform. This also includes the capability for software development, machine learning, and artificial intelligence development, including Python [69], Matrix Laboratory [70], and a suite of tools within Jupyter Notebooks (Sagemaker instance) [71] such as TensorFlow. The environment is also being enhanced to support multi-omic data [72] analysis through pipelines, using tools such as Plink [73] and Nexflow [74] with resource scheduling through Amazon Web Services Batch [75]. The analytical platform provides graphics processing units and high-performance computing capabilities.

For larger projects, where the number of rows is too high to manage in other formats, the HIC Safe Haven provides the data in an RDBMS in the analytical platform for use by researchers.

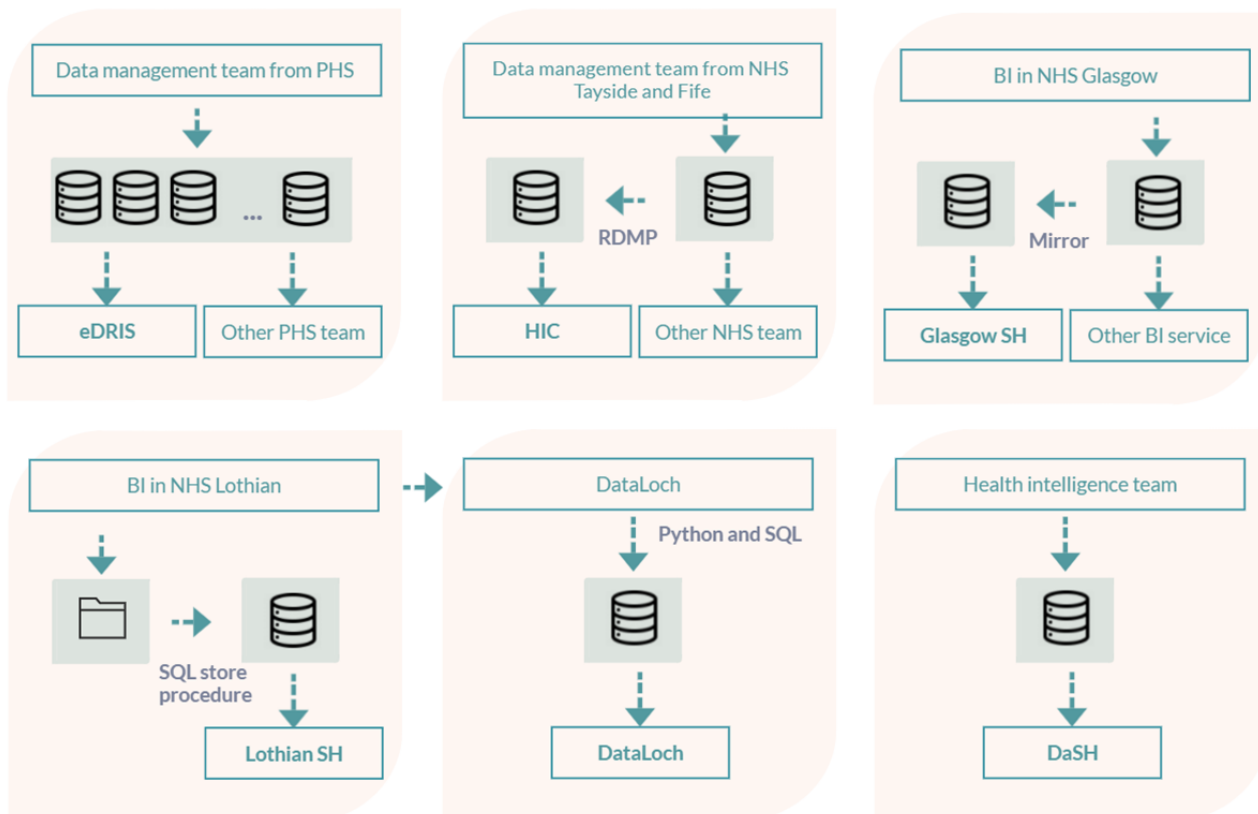
Researchers rely upon Safe Havens to archive the raw data and derived data products from their analysis as they are not permitted to export any of that data from an analytical platform. A research project may be archived for a period of between 5 and 30 years, depending on regulations and researcher or funder requirements. Archiving typically takes place using the analytical platform infrastructure. There can be significant costs for storing and securing large amounts of data, and a policy for long-term archival is being jointly developed by the Scottish Safe Havens.

Data Repository Infrastructure

Safe Havens have their source EHRs on the NHS network, which are transferred to the service network (where the cohort building and linkage take place; Figure 1) when required. They create cohorts and associated data on the NHS infrastructure before the data go through the Safe Haven functions of linkage, deidentification, and transfer to analytical platform or platforms for researchers to access. The exception is eDRIS, which has some data sets managed securely in a university environment by EPCC.

The infrastructure and the ETL process for those data repositories vary among Safe Havens (Figure 4).

Figure 4. Safe Haven data repository networks. Upper row from left to right: electronic Data Research and Innovation Service, Health Informatics Center, and Glasgow Safe Haven. The lower row from left to right: Lothian, DataLoch, and Grampian Data Safe Haven Safe Haven. BI: Business Intelligence and Informatics; DaSH: Grampian Data Safe Haven; eDRIS: electronic Data Research and Innovation Service; HIC: Health Informatics Center; NHS: National Health Service; PHS: Public Health Scotland; RDMP: Research Data Management Platform; SH: Safe Haven.



As shown in Table 1, eDRIS has access to 85 national NHS data sets; these are updated and maintained by PHS. There are data sets that eDRIS cannot access routinely; however, for a known cohort, they can request data from other teams within PHS. The data management team within PHS performs quality assurance after ETL using R or SPSS (in cases of legacy data). In addition to providing data to eDRIS, the data are also used to run hundreds of different reports and publications by other teams within PHS.

HIC’s data repository infrastructure and NHS Tayside or Fife data are colocated within the same data center. The HIC runs the University of Dundee–owned and managed servers connected to the NHS Tayside network and receives regular feeds of data from the NHS Tayside clinical systems and PHS (covering consented cohorts of research data and for the patients within the Tayside and Fife regions). The RDMP tool takes data from the feeds and performs ETL to clean and transform the data, which are then stored within structured databases.

Glasgow Safe Haven’s data repository mirrors some data from the routine data systems that are maintained by Business Intelligence and Informatics in NHS Glasgow. For custom NHS data or data collected for research projects (eg, some SMR, Prescribing Information System, all audit data, device data, and trial data), Glasgow Safe Haven staff will conduct the ETL themselves.

Lothian or DataLoch data repositories residing on the NHS Lothian information technology infrastructure use stored Python or SQL to load data updates from PHS and data feeds via

Business Intelligence and Informatics within NHS Lothian for copies of data from local clinical systems.

NHS Grampian’s health intelligence team updates the DaSH Safe Haven repository monthly. Both Lothian or DataLoch and DaSH Safe Haven deal with changing data formats by separating new and old data.

Discussion

Principal Findings

Overview

Scotland has many strengths regarding enabling EHRs for reuse. There is a single NHS where patients are allocated CHI numbers that can be used to link their entire patient history. The Scottish network of Safe Havens has similar architectures, adheres to the Scottish Safe Haven Charter [6], is accredited by the Scottish Government and International Organization for Standardization 27001 [55,56], and is the common information security standard. Each regional Safe Haven has a rich and deep data source from their local health boards, and the national Safe Haven has the breadth of a whole-population view and close links to other health and social care data sources.

All the Safe Havens make use of two networks: (1) an analytical platform set up within university-managed networks and (2) data repositories set up mainly on NHS networks. The existing operational separation of source data repository, linkage infrastructure, and analytical platform provides a solid

foundation for increasing collaborative work across national and multi-Safe Haven projects.

There are some barriers, as highlighted in the *Scottish Safe Haven* section, to making multi-Safe Haven projects as streamlined as possible. Addressing them in a coordinated manner would pave the way to achieving a federated system of Safe Havens in Scotland. These opportunities for improvement are described in the following sections.

Data Visibility

The depth of the Scottish data, which are hosted by regional Safe Havens (described in the *Scottish NHS Data Sources* section), is not widely used by the wider community. These data sets are unique to each regional Safe Haven and are difficult to bring up to a consistent national level. Interactions with researchers for feasibility, generating aggregate numbers, scoping projects, and providing quotes for work can be resource intensive. Many data requests to regional Safe Havens are from frequent users who know the specific data structures and terminologies used by each Safe Haven well. In making the regional data more visible and accessible, researchers will be better able to run projects using data from multiple Safe Havens.

Data Standards and Common Data Models

As shown in [Table 1](#), the Safe Havens accept data that use any number of standards. Owing to the processing efficiency, the *create and destroy* model mandated by the Safe Haven Charter, and the fact that researchers normally prefer to have the original data, there has been little attempt to harmonize extracted data for placement in analytical platforms. If common data models such as Observational Medical Outcomes Partnership [76] and i2b2 star schema [77] were used, either for data repositories or analytical platforms, the burden on multi-Safe Haven projects would be reduced, and operational access to data would be faster and more predictable.

Governance

In the Safe Haven network, access to linked data is fragmented, with researchers and health care providers having to work with Safe Havens to obtain local, regional, or national data controllers' approvals. Data governance, in general, is much easier at a local level. At the Scottish national level, application forms for submission to the Public Benefit and Privacy Panel for Health and Social Care and Statistics Public Benefit and Privacy Panel are normally required. This is a complex process and can take significant time for review and approval.

With the experience and knowledge gained from supporting projects requiring diverse local and regional data sets [78,79], and building capability for a federated network, we propose that the following aspects of the network be addressed in future research:

- The establishment of a shared method for cataloging and managing metadata would facilitate data discovery and research feasibility.
- To facilitate cross-Safe Haven data governance, standardization of the application interface specifications to Safe Havens would permit easier cross-access of Safe Havens by researchers.
- Health care delivery is explicitly devolved to local structures via health and social care partnerships in Scotland and associated legislation. With functions devolved to individual health boards, the linking of regional, available data will require greater collaboration across the organizations and appropriate benefit shares.

Conclusions

The Safe Haven network in Scotland has supported over a thousand projects in the past 5 years, underpinning world-class research outputs. It not only brought grant research, jobs, and funding to Scotland but also enabled international health research with many countries such as Brazil and India.

This paper reports on the operational assessment of each of the 4 regional Safe Havens and the national Safe Haven. We compared a set of functions and services related to data forming part of EHRs in Scotland. We have described the operation of Scottish Safe Haven data services and functions and their technical implementation from the following points of view: (1) data governance and workflow, (2) data discovery and metadata, (3) data linkage, (4) data deidentification, (5) analytical platforms, and (6) data repository infrastructure. The results obtained should assist the Scottish Safe Havens to scale operations to larger cohorts and more diverse data, reduce timescales and operate more cost-effectively. More importantly, this work identified the responsibilities and work needed for each Scottish Safe Haven to contribute to the building of a national federated data-sharing platform. Although this paper has focused on experiences across Scotland, the findings will be of interest nationally or internationally to inform the understanding of the challenges that exist for the reuse of EHR data in clinical and other kinds of research.

Acknowledgments

This work was supported by Health Data Research United Kingdom (636000/RA4624 and 636000/RA4707), which receives its funding from Health Data Research United Kingdom Ltd (HDR-5012), funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government, Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and the Wellcome Trust. The work was supported by the Scottish Government Chief Scientist Office through the Safe Haven Metadata project.

Conflicts of Interest

None declared.

References

1. World Health Organization. From Innovation to Implementation: eHealth in the WHO European Region. Geneva: World Health Organization - Regional Office for Europe; 2016.
2. Geissbuhler A, Safran C, Buchan I, Bellazzi R, Labkoff S, Eilenberg K, et al. Trustworthy reuse of health data: a transnational perspective. *Int J Med Inform* 2013 Jan;82(1):1-9. [doi: [10.1016/j.ijmedinf.2012.11.003](https://doi.org/10.1016/j.ijmedinf.2012.11.003)] [Medline: [23182430](https://pubmed.ncbi.nlm.nih.gov/23182430/)]
3. Doiron D, Raina P, Fortier I, Linkage Between Cohorts and Health Care Utilization Data: Meeting of Canadian Stakeholders workshop participants. Linking Canadian population health data: maximizing the potential of cohort and administrative data. *Can J Public Health* 2013 Mar 06;104(3):258-261 [FREE Full text] [doi: [10.17269/cjph.104.3775](https://doi.org/10.17269/cjph.104.3775)] [Medline: [23823892](https://pubmed.ncbi.nlm.nih.gov/23823892/)]
4. A matter of life and death: how your health information can make a difference. AMRC. URL: <https://www.amrc.org.uk/Handlers/Download.ashx?IDMF=652f0e60-315b-4258-8e81-fe1fdf8b65dc> [accessed 2022-01-17]
5. Lea NC, Nicholls J, Dobbs C, Sethi N, Cunningham J, Ainsworth J, et al. Data safe havens and trust: toward a common understanding of trusted research platforms for governing secure and ethical health research. *JMIR Med Inform* 2016 Jun 21;4(2):e22 [FREE Full text] [doi: [10.2196/medinform.5571](https://doi.org/10.2196/medinform.5571)] [Medline: [27329087](https://pubmed.ncbi.nlm.nih.gov/27329087/)]
6. Charter for safe havens in Scotland: handling unconsented data from national health service patient records to support research and statistics. Scottish Government. URL: <https://tinyurl.com/2p97v7fz> [accessed 2022-01-17]
7. iCAIRD. URL: <https://icaird.com/> [accessed 2022-01-17]
8. Research-Data-Scotland. URL: <https://www.researchdata.scot/> [accessed 2022-01-17]
9. Caldicott F. Information: to share or not to share? The Information Governance Toolkit. 2013. URL: <https://tinyurl.com/ybxmutar> [accessed 2022-01-17]
10. Thomas R, Walport M. Data sharing review report. Data Sharing Review. 2008. URL: <https://amberhawk.typepad.com/files/thomas-walport-datasharingreview2008.pdf> [accessed 2022-01-17]
11. Burton PR, Murtagh MJ, Boyd A, Williams JB, Dove ES, Wallace SE, et al. Data Safe Havens in health research and healthcare. *Bioinformatics* 2015 Oct 15;31(20):3241-3248 [FREE Full text] [doi: [10.1093/bioinformatics/btv279](https://doi.org/10.1093/bioinformatics/btv279)] [Medline: [26112289](https://pubmed.ncbi.nlm.nih.gov/26112289/)]
12. Jones KH, Ford DV, Thompson S, Lyons R. A profile of the SAIL Databank on the UK Secure Research Platform. *Int J Popul Data Sci* 2019 Nov 20;4(2):1134 [FREE Full text] [doi: [10.23889/ijpds.v4i2.1134](https://doi.org/10.23889/ijpds.v4i2.1134)] [Medline: [34095541](https://pubmed.ncbi.nlm.nih.gov/34095541/)]
13. Jones KH, Ford DV, Ellwood-Thompson S, Lyons RA. The UK Secure eResearch Platform for public health research: a case study. *Lancet* 2016 Nov;388:S62. [doi: [10.1016/s0140-6736\(16\)32298-x](https://doi.org/10.1016/s0140-6736(16)32298-x)]
14. Trusted Research Environments (TRE): a strategy to build public trust and meet changing health data science needs. HDRUK. URL: https://ukhealthdata.org/wp-content/uploads/2020/07/200723-Alliance-Board_Paper-E_TRE-Green-Paper.pdf [accessed 2022-01-17]
15. Sciensano. URL: <https://www.sciensano.be/en> [accessed 2022-01-17]
16. Joseph K, Mahey J. Validation of perinatal data in the Discharge Abstract Database of the Canadian Institute for Health Information. *Chronic Dis Can* 2009 May;29(3):96-101. [doi: [10.24095/hpcdp.29.3.01](https://doi.org/10.24095/hpcdp.29.3.01)]
17. Schneider M, Radbone CG, Vasquez SA, Palfy M, Stanley AK. Population data centre profile: SA NT DataLink (South Australia and Northern Territory). *Int J Popul Data Sci* 2019 Dec 05;4(2):1136 [FREE Full text] [doi: [10.23889/ijpds.v4i2.1136](https://doi.org/10.23889/ijpds.v4i2.1136)] [Medline: [32935038](https://pubmed.ncbi.nlm.nih.gov/32935038/)]
18. Kavianpour S, Sutherland J, Mansouri-Benssassi E, Coull N, Jefferson E. A review of trusted research environments to support next generation capabilities based on interview analysis. *JMIR Preprints* 2021:33720. [doi: [10.2196/preprints.33720](https://doi.org/10.2196/preprints.33720)]
19. Nind T, Galloway J, McAllister G, Scobbie D, Bonney W, Hall C, et al. The research data management platform (RDMP): a novel, process driven, open-source tool for the management of longitudinal cohorts of clinical data. *Gigascience* 2018 Jul 01;7(7):giy060 [FREE Full text] [doi: [10.1093/gigascience/giy060](https://doi.org/10.1093/gigascience/giy060)] [Medline: [29790950](https://pubmed.ncbi.nlm.nih.gov/29790950/)]
20. Jefferson ER, Truccob E. The challenges of assembling, maintaining and making available large data sets of clinical data for research. In: *Computational Retinal Image Analysis*. Amsterdam: Elsevier; 2019:429-444.
21. Joined-up data for better decisions: guiding principles for data linkage. Scottish Government. 2012. URL: <https://www.gov.scot/publications/joined-up-data-better-decisions-guiding-principles-data-linkage/> [accessed 2022-01-17]
22. Chief Scientist Office. URL: <https://www.cso.scot.nhs.uk/> [accessed 2022-01-17]
23. Products and Services. eDRIS. URL: <https://www.isdscotland.org/Products-and-Services/eDRIS/> [accessed 2022-01-17]
24. COVID-19 pandemic response. Public Health Scotland. URL: <https://www.isdscotland.org/Products-and-Services/eDRIS/> [accessed 2022-01-17]
25. EPCC. URL: <https://www.epcc.ed.ac.uk/> [accessed 2022-01-17]
26. Grampian Population Platform: datasets permissioned for access for research purposes via DaSH. DaSH. URL: <https://www.abdn.ac.uk/iahs/facilities/dash-datasets.php> [accessed 2022-01-17]
27. Health Informatics Centre - trusted research environment. HIC. URL: <https://www.dundee.ac.uk/hic> [accessed 2022-01-17]

28. NHS-Greater-Glasgow-and-Clyde. Glasgow Safe Haven Services. URL: <https://www.nhsggc.org.uk/about-us/professional-support-sites/glasgow-safe-haven/services/> [accessed 2022-01-17]
29. Datasets. Lothian-Research-Safe-Haven. URL: <http://accord.scot/researcher-access-research-data-nrs-safe-haven/datasets> [accessed 2022-01-17]
30. COVID-19 collaborative. DATALOCH. URL: <https://www.ed.ac.uk/usher/dataloch/covid-19-collaborative> [accessed 2022-01-17]
31. Data linkage. eDRIS. URL: <https://www.isdscotland.org/Products-and-Services/eDRIS/FAQ-eDRIS/#e1> [accessed 2022-01-17]
32. RDS - Research Data Scotland. URL: <https://www.researchdata.scot/> [accessed 2022-01-17]
33. National datasets. National Data Catalogue. URL: <https://www.ndc.scot.nhs.uk/National-Datasets/> [accessed 2022-01-17]
34. ePharmacy user guide (Scotland). INPS. URL: <https://tinyurl.com/339xhbet> [accessed 2022-01-17]
35. Glossary of terms: practice level prescribing data. Information Services Division, Scotland. 2018. URL: https://www.isdscotland.org/health-topics/prescribing-and-medicines/docs/Open_Data_Glossary_of_Terms.pdf?1 [accessed 2022-01-17]
36. Alvarez-Madrado S, McTaggart S, Nangle C, Nicholson E, Bennie M. Data resource profile: The Scottish National Prescribing Information System (PIS). *Int J Epidemiol* 2016 Jun 10;45(3):714-715 [FREE Full text] [doi: [10.1093/ije/dyw060](https://doi.org/10.1093/ije/dyw060)] [Medline: [27165758](https://pubmed.ncbi.nlm.nih.gov/27165758/)]
37. CHI Number. ISD Scotland Data Dictionary. URL: <https://tinyurl.com/3bj3xr5e> [accessed 2022-01-17]
38. Generation Scotland. URL: <https://www.ed.ac.uk/generation-scotland> [accessed 2022-01-17]
39. McKinstry B, Sullivan FM, Vasishta S, Armstrong R, Hanley J, Haughney J, et al. Cohort profile: the Scottish Research register SHARE. A register of people interested in research participation linked to NHS data sets. *BMJ Open* 2017 Feb 01;7(2):e013351 [FREE Full text] [doi: [10.1136/bmjopen-2016-013351](https://doi.org/10.1136/bmjopen-2016-013351)] [Medline: [28148535](https://pubmed.ncbi.nlm.nih.gov/28148535/)]
40. EMBARC - The European Bronchiectasis Registry. URL: <https://www.bronchiectasis.eu/registry> [accessed 2022-01-17]
41. DARTS (Diabetes Audit and Research in Tayside Scotland). GoDARTS. URL: <https://godarts.org/> [accessed 2022-01-17]
42. Childsmile – Improving the Oral Health of Children in Scotland. URL: <http://www.child-smile.org.uk/> [accessed 2022-01-17]
43. Sawhney S, Marks A, Fluck N, Levin A, Prescott G, Black C. Intermediate and long-term outcomes of survivors of acute kidney injury episodes: a large population-based cohort study. *Am J Kidney Dis* 2017 Jan;69(1):18-28 [FREE Full text] [doi: [10.1053/j.ajkd.2016.05.018](https://doi.org/10.1053/j.ajkd.2016.05.018)] [Medline: [27555107](https://pubmed.ncbi.nlm.nih.gov/27555107/)]
44. Ayorinde AA, Wilde K, Lemon J, Campbell D, Bhattacharya S. Data resource profile: The Aberdeen Maternity and Neonatal Databank (AMND). *Int J Epidemiol* 2016 Apr 22;45(2):389-394. [doi: [10.1093/ije/dyv356](https://doi.org/10.1093/ije/dyv356)] [Medline: [26800750](https://pubmed.ncbi.nlm.nih.gov/26800750/)]
45. Children of the 1950s. Aberdeen Birth Cohorts. URL: <https://www.abdn.ac.uk/birth-cohorts/1950s/> [accessed 2022-01-17]
46. SCI-Diabetes. Scottish Care Information. URL: <https://www.sci-diabetes.scot.nhs.uk/> [accessed 2022-01-17]
47. Denaxas S, Gonzalez-Izquierdo A, Direk K, Fitzpatrick NK, Fatemifar G, Banerjee A, et al. UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *J Am Med Inform Assoc* 2019 Dec 01;26(12):1545-1559 [FREE Full text] [doi: [10.1093/jamia/ocz105](https://doi.org/10.1093/jamia/ocz105)] [Medline: [31329239](https://pubmed.ncbi.nlm.nih.gov/31329239/)]
48. International statistical classification of diseases and related health problems (ICD). World Health Organization. URL: <https://www.who.int/standards/classifications/classification-of-diseases> [accessed 2022-01-17]
49. The National Data Catalogue (NDC). Public Health Scotland. URL: <https://www.ndc.scot.nhs.uk/> [accessed 2022-01-17]
50. Dataset inventory. HIC - University of Dundee. URL: <https://www.dundee.ac.uk/hic/datalinkageservice/datasetinventory/> [accessed 2022-01-17]
51. Specification for phase 2: technology partnership. Health Data Research Innovation Gateway, UK. 2019. URL: <https://tinyurl.com/5n72htw7> [accessed 2022-01-17]
52. Trust data provenance. Centre for Health Data Science. URL: <https://www.abdn.ac.uk/achds/research/data-provenance-172.php> [accessed 2022-01-17]
53. SMR crib sheets. Public Health Scotland. URL: <https://www.ndc.scot.nhs.uk/Data-Dictionary/SMR-Crib-Sheets/> [accessed 2022-01-17]
54. Scottish Index of Multiple Deprivation 2020. Scottish Government. URL: <https://tinyurl.com/2p9cd7t4> [accessed 2022-01-17]
55. Calder A, Watkins SG. Information Security Risk Management for ISO 27001/ISO 27002, Third Edition. Cambridgeshire, UK: IT Governance Publishing; 2019.
56. Information technology - Security techniques - Information security management systems - Requirements. International-Organization-for-Standardization, ISO/IEC 27001. 2013. URL: <https://synergia.org/wp-content/uploads/2021/02/ISO-27001-2013-Information-Security-Management-EN.pdf> [accessed 2022-01-17]
57. SWEENEY L. k-anonymity: a model for protecting privacy. *Int J Uncertain Fuzz Knowl Based Syst* 2012 May 02;10(05):557-570. [doi: [10.1142/s0218488502001648](https://doi.org/10.1142/s0218488502001648)]
58. Machanavajjhala A, Kifer D, Gehrke J, Venkitasubramaniam M. L-diversity: privacy beyond k-anonymity. *ACM Trans Knowl Discov Data* 2007 Mar;1(1):3. [doi: [10.1145/1217299.1217302](https://doi.org/10.1145/1217299.1217302)]
59. Samarati P. Protecting respondents identities in microdata release. *IEEE Trans Knowl Data Eng* 2001;13(6):1010-1027. [doi: [10.1109/69.971193](https://doi.org/10.1109/69.971193)]
60. Scottish Medical Imaging (SMI) Service. Public Health Scotland. URL: <https://beta.isdscotland.org/products-and-services/scottish-medical-imaging-smi-service/> [accessed 2022-01-17]

61. Nind T, Sutherland J, McAllister G, Hardy D, Hume A, MacLeod R, et al. An extensible big data software architecture managing a research resource of real-world clinical radiology data linked to other health data from the whole Scottish population. *Gigascience* 2020 Sep 29;9(10):giaa095 [FREE Full text] [doi: [10.1093/gigascience/giaa095](https://doi.org/10.1093/gigascience/giaa095)] [Medline: [32990744](https://pubmed.ncbi.nlm.nih.gov/32990744/)]
62. Jackson R, Kartoglu I, Stringer C, Gorrell G, Roberts A, Song X, et al. CogStack - experiences of deploying integrated information retrieval and extraction services in a large National Health Service Foundation Trust hospital. *BMC Med Inform Decis Mak* 2018 Jun 25;18(1):47 [FREE Full text] [doi: [10.1186/s12911-018-0623-9](https://doi.org/10.1186/s12911-018-0623-9)] [Medline: [29941004](https://pubmed.ncbi.nlm.nih.gov/29941004/)]
63. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004 Dec;57(12):1288-1294. [doi: [10.1016/j.jclinepi.2004.03.012](https://doi.org/10.1016/j.jclinepi.2004.03.012)] [Medline: [15617955](https://pubmed.ncbi.nlm.nih.gov/15617955/)]
64. Tonelli M, Wiebe N, Fortin M, Guthrie B, Hemmelgarn BR, James MT, Alberta Kidney Disease Network. Methods for identifying 30 chronic conditions: application to administrative data. *BMC Med Inform Decis Mak* 2015 Apr 17;15(1):31 [FREE Full text] [doi: [10.1186/s12911-015-0155-5](https://doi.org/10.1186/s12911-015-0155-5)] [Medline: [25886580](https://pubmed.ncbi.nlm.nih.gov/25886580/)]
65. IBM SPSS Statistics. IBM. URL: <https://www.ibm.com/uk-en/products/spss-statistics> [accessed 2022-01-17]
66. STATA. URL: <https://www.stata.com/> [accessed 2022-01-17]
67. SAS - Analytics Software & Solutions. URL: https://www.sas.com/en_gb/home.html [accessed 2022-01-17]
68. The R Project for Statistical Computing. The R Foundation. URL: <https://www.r-project.org/> [accessed 2022-01-17]
69. Python. URL: <https://www.python.org/> [accessed 2022-01-17]
70. MATLAB. URL: <https://www.mathworks.com/products/matlab.html> [accessed 2022-01-17]
71. Use Amazon SageMaker Notebook Instances. Amazon. URL: <https://docs.aws.amazon.com/sagemaker/latest/dg/nbi.html> [accessed 2022-01-17]
72. Palsson B, Zengler K. The challenges of integrating multi-omic data sets. *Nat Chem Biol* 2010 Nov 18;6(11):787-789. [doi: [10.1038/nchembio.462](https://doi.org/10.1038/nchembio.462)] [Medline: [20976870](https://pubmed.ncbi.nlm.nih.gov/20976870/)]
73. Plink. URL: <https://zzz.bwh.harvard.edu/plink/> [accessed 2022-01-17]
74. Nextflow. URL: <https://www.nextflow.io/> [accessed 2022-01-17]
75. AWS Batch. Amazon. URL: <https://aws.amazon.com/batch/> [accessed 2022-01-17]
76. OMOP Common Data Model. OHDSI. URL: <https://www.ohdsi.org/data-standardization/the-common-data-model/> [accessed 2022-01-17]
77. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). *J Am Med Inform Assoc* 2010 Feb 26;17(2):124-130 [FREE Full text] [doi: [10.1136/jamia.2009.000893](https://doi.org/10.1136/jamia.2009.000893)] [Medline: [20190053](https://pubmed.ncbi.nlm.nih.gov/20190053/)]
78. Shah AS, Anand A, Strachan FE, Ferry AV, Lee KK, Chapman AR, et al. High-sensitivity troponin in the evaluation of patients with suspected acute coronary syndrome: a stepped-wedge, cluster-randomised controlled trial. *Lancet* 2018 Sep;392(10151):919-928. [doi: [10.1016/s0140-6736\(18\)31923-8](https://doi.org/10.1016/s0140-6736(18)31923-8)]
79. No authors listed. High sensitivity cardiac troponin and the under-diagnosis of myocardial infarction in women: prospective cohort study. *Br Med J* 2015 Feb 03;350:h626 [FREE Full text] [doi: [10.1136/bmj.h626](https://doi.org/10.1136/bmj.h626)] [Medline: [25646958](https://pubmed.ncbi.nlm.nih.gov/25646958/)]

Abbreviations

- CHI:** community health index
- CHILI:** community health index Linkage Team
- DaSH:** Grampian Data Safe Haven
- eDRIS:** electronic Data Research and Innovation Service
- EHR:** electronic health record
- EPCC:** Edinburgh Parallel Computing Centre
- ETL:** extract, transform, and load
- HIC:** Health Informatics Center
- ICD:** International Statistical Classification of Diseases and Related Health Problems
- NHS:** National Health Service
- NRS:** National Records Scotland
- PHS:** Public Health Scotland
- RDBMS:** Relational Database Management System
- RDMP:** Research Data Management Platform
- SCI-Diabetes:** Scottish Care Information–Diabetes
- SMR:** Scottish Morbidity Record

Edited by R Kukafka; submitted 09.07.21; peer-reviewed by I Shubina, I Schiering; comments to author 05.10.21; revised version received 18.11.21; accepted 03.12.21; published 09.03.22.

Please cite as:

Gao C, McGilchrist M, Mumtaz S, Hall C, Anderson LA, Zurowski J, Gordon S, Lumsden J, Munro V, Wozniak A, Sibley M, Banks C, Duncan C, Linksted P, Hume A, Stables CL, Mayor C, Caldwell J, Wilde K, Cole C, Jefferson E

A National Network of Safe Havens: Scottish Perspective

J Med Internet Res 2022;24(3):e31684

URL: <https://www.jmir.org/2022/3/e31684>

doi: [10.2196/31684](https://doi.org/10.2196/31684)

PMID: [35262495](https://pubmed.ncbi.nlm.nih.gov/35262495/)

©Chuang Gao, Mark McGilchrist, Shahzad Mumtaz, Christopher Hall, Lesley Ann Anderson, John Zurowski, Sharon Gordon, Joanne Lumsden, Vicky Munro, Artur Wozniak, Michael Sibley, Christopher Banks, Chris Duncan, Pamela Linksted, Alastair Hume, Catherine L Stables, Charlie Mayor, Jacqueline Caldwell, Katie Wilde, Christian Cole, Emily Jefferson. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 09.03.2022. 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.

Viewpoint

Scale-up of Digital Innovations in Health Care: Expert Commentary on Enablers and Barriers

Hannes Schlieter¹, PhD; Lisa A Marsch², PhD; Diane Whitehouse³, MSc; Lena Otto¹, PhD; Ana Rita Londral^{4,5}, PhD; Gisbert Wilhelm Teepe⁶, MSc; Martin Benedict¹, PhD; Joseph Ollier⁶, MSc; Tom Ulmer⁷, Mag; Nathalie Gasser⁸, MSc; Sabine Ultsch⁹, BSc; Bastian Wollschlaeger¹, Dipl-Ing; Tobias Kowatsch^{6,9}, PhD

¹Research Group Digital Health, Faculty of Business and Economics, Technische Universität Dresden, Dresden, Germany

²Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States

³European Health Telematics Association, Brussels, Belgium

⁴Value for Health CoLAB, NOVA Medical School, Nova University of Lisbon, Lisbon, Portugal

⁵Comprehensive Health Research Center, NOVA Medical School, Nova University of Lisbon, Lisbon, Portugal

⁶Centre for Digital Health Interventions, Department of Management, Technology and Economics, ETH Zurich, Zurich, Switzerland

⁷Institute of Information and Process Management, Eastern Switzerland University of Applied Sciences, St. Gallen, Switzerland

⁸Swiss Lung Association, Bern, Switzerland

⁹Centre for Digital Health Interventions, Institute of Technology Management, University of St. Gallen, St. Gallen, Switzerland

Corresponding Author:

Tobias Kowatsch, PhD

Centre for Digital Health Interventions

Department of Management, Technology and Economics

ETH Zurich

Weinbergstrasse 56/58

Zurich, 8092

Switzerland

Phone: 41 44 632 94 8

Email: tobias.kowatsch@unig.ch

Abstract

Health care delivery is undergoing a rapid change from traditional processes toward the use of digital health interventions and personalized medicine. This movement has been accelerated by the COVID-19 crisis as a response to the need to guarantee access to health care services while reducing the risk of contagion. Digital health scale-up is now also vital to achieve population-wide impact: it will only accomplish sustainable effects if and when deployed into regular health care delivery services. The question of how sustainable digital health scale-up can be successfully achieved has, however, not yet been sufficiently resolved. This paper identifies and discusses enablers and barriers for scaling up digital health innovations. The results discussed in this paper were gathered by scientists and representatives of public bodies as well as patient organizations at an international workshop on scaling up digital health innovations. Results are explored in the context of prior research and implications for future work in achieving large-scale implementations that will benefit the population as a whole.

(*J Med Internet Res* 2022;24(3):e24582) doi:[10.2196/24582](https://doi.org/10.2196/24582)

KEYWORDS

digital health; health care delivery; health interventions; digital health services; enablers; barriers

Introduction

Health care delivery is undergoing a rapid change from traditional processes toward the use of digital health services [1-3], that is, “tools and services that use information and communication technologies to improve prevention, diagnosis, treatment, monitoring, and management of health-related issues

and to monitor and manage lifestyle habits that impact health” [4]. With this shift also comes a move to precision medicine [5] and precision health [6]. Hospitals and health care providers introduce hospital information systems [7,8], electronic health records [9-11], and telemedicine solutions for more efficient workflows within and beyond institutions [12,13]. Many people are choosing among a wide range of digital health services provided by wearables and mobile phone apps that support their

self-management, health, and well-being [14]. These technologies may increasingly employ digital biomarkers to sense states of vulnerability [15,16], text- or voice-based conversational agents for intervention delivery [17-19], or a mixture of human and digital support via blended treatments [20,21]. These digital health services may be able to intervene with the right type of support, at the right time, while including contextual factors that offer a distinct contribution outside of human-delivered care [22]. Even though the number of existing services is growing, not many users are currently taking full advantage of these services. However, a better adoption by users would facilitate the diffusion of these innovative services [23]. The question of how sustainable digital health diffusion can be successfully achieved by scaling up individual services, that is, reaching more people benefitting from them [24], is not sufficiently solved yet, despite plenty of theoretical insights being available. To also consider a practical perspective, experts of digital health services gathered at a workshop. The results of the workshop combined practice- and research-based perspectives, which were matched with theoretical insights in this paper. Therefore, this paper goes one step further toward proposing a solution for the issue of scaling up digital health innovations by identifying and discussing barriers and enablers from a holistic point of view. Owing to the special circumstances of the workshop being held right at the onset of the COVID-19 pandemic in Europe, a discussion on how these barriers and enablers were affected by the COVID-19 crisis is also provided.

Workshop Context and Contents

Barriers and enablers for scaling up digital health innovations were identified and discussed in the context of a conference workshop. The workshop participants, who are also coauthors of this paper, were both junior (n=8) and senior scientists (n=2) and representatives from nongovernmental organizations (n=2) and a home care provider (n=1). The participants came from diverse countries with backgrounds in public health, implementation science, information systems research, and computer science. Many of the participants had several years of experience with the design and implementation of digital health services. All participants came together at the “1st International Workshop on Best Practices for Scaling-Up Digital Innovations in Health care–Scale-IT-up!” The workshop was held at the BIOSTEC conference in Valletta, Malta, on February 25, 2020 and spanned 3 sessions, with 5 papers presented by some of the workshop’s participants [25-29]. Two keynote speeches on digital health innovations were given by Lisa A Marsch from the Dartmouth College in the United States and Diane Whitehouse from the European Health Telematics Association in Belgium, who both have extensive experience in scaling up digital health innovations. For example, Dr Marsch codeveloped the most empirically supported digital behavioral therapy for substance use disorders: it became the very first “prescription digital therapeutic” approved in the United States by the Food and Drug Administration [30-32]. Diane Whitehouse, as principal eHealth consultant in European Health Telematics Association, has followed a range of scaling-up projects. Examples include one related to telemedicine [33] and another related to integrated care [34]. Based on this experience, both keynotes presented insights from various international

initiatives and projects. With the agreement of the speakers, the presentations and keynotes have been made available to the general public [35]. Each presentation provided a different focus on what drives the successful scale-up of digital health innovations: for example, how financial incentives need to be defined to motivate patients to adopt digital health innovations successfully. Based on the input of all presenters, the last workshop session featured a discussion of all 13 speakers and workshop participants on best practices and challenges for scaling up digital health innovations. This last workshop session forms the basis for the method used to identify the relevant set of enablers and barriers.

Methods

The discussion during the last session of the workshop followed a structured approach. To this end, 2 topic leaders—one each for the topics of barriers and enablers—were determined. The remaining participants were then split into 2 groups. Both groups undertook a 2-round group process, with discussions on barriers and enablers.

In the first round, the groups identified either enablers or barriers for scaling up digital health innovations depending on their initial topic assignment. This identification was carried out according to the brainwriting technique [36], with each participant writing down items individually. This process enabled the participants to take their time and to be equally involved in the process. Afterwards, each participant presented and explained their list of enablers or barriers to the group so that the ideas could be consolidated and clustered. This process took 20 minutes before group members switched to the other topic (ie, from enablers to barriers and vice versa). The 2 topic leaders remained to inform the other group members about the intermediate results.

In the second round, the other group was informed about the results of the first round and could extend and revise these findings. Finally, all participants were given the final results, which were discussed and consolidated until a group consensus between all 13 participants was reached. Owing to our clustering, some items could have been mentioned multiple times and therefore carried more weight than others. However, no weights were added after identifying and discussing the findings. Afterwards, the raw results of each topic were digitized and categorized, and duplicates were removed. Finally, the results were aggregated and aligned with the existing work of DeLone and McLean [37] regarding enablers and Kowatsch et al [3] regarding both enablers and barriers. This allowed us to validate the results gained and to assess whether new aspects have been identified.

Results

In total, 36 enablers and 33 barriers were identified in the workshop session. To align these enablers and barriers with prior research, they were grouped in categories classified by DeLone and McLean [37] as enablers and Kowatsch et al [3] as both enablers and barriers. To further understand the context of digital innovations in health care, the aspects were also

grouped according to 5 various levels of influence [38], namely, the micro, meso, macro, and the technology/innovation level, or in an overarching category. Some barriers and enablers cannot be influenced by only one of the 4 levels and were therefore grouped into an overarching category.

Overarching aspects that are influenced by all stakeholders involved on all levels are leadership, culture, interdisciplinary cocreation, innovation characteristics, and methodology. Although missing leadership in projects is an exemplary barrier hindering the scale-up of digital health innovations, dialogue between all stakeholders involved can be supportive. Further, a culture in favor of digital health innovations needs to exist, as it otherwise hinders the scale-up process when not existent.

If interdisciplinary cocreation is missing, this can be a barrier, while collaboration between stakeholders or stakeholder engagement are exemplary enablers: indeed, the importance of the “quadruple helix” [39] and engagement cannot be understated. Characteristics of the innovation itself and the methodology mainly act as barriers for digital health innovations. Participants referred to a lack of trust when transferring existing solutions to new contexts (the “not invented here” dilemma). Named as a further potential barrier, a pace of multistakeholder innovation that is too high may lead to piecemeal approaches or to not paying heed to past successes/failures in other settings when designing new approaches. All overarching barriers and enablers are presented in Table 1.

Table 1. Overarching enablers and barriers for scaling up digital health innovations identified in the workshop and aligned with prior work.

Category	Enabler	Barrier
Leadership ^a	<p>Continuous dialogue between academia, industry, government, and other stakeholders to facilitate policy-relevant research and increase scale-up of science-based best practices</p> <p>Visionary leadership: clear idea of leadership what the digital health innovation should look like in the future</p> <p>Care management: managing medical conditions more effectively by patient-centered approach that is designed to assist patients and their support system</p>	Missing leadership in projects
User culture ^b	— ^c	Inherent characteristics and preferences of specific user groups
Interdisciplinary cocreation ^a	<p>Collaboration between medical experts, computer scientists, business experts, etc</p> <p>Continuous dialogue between academia, industry, government, and other stakeholders to facilitate policy-relevant research and increase scale-up of scientifically validated best practices</p> <p>Employee involvement (direct participation of staff, eg, applying own ideas, expertise, efforts for developing digital health interventions)</p> <p>Engagement of diverse stakeholders/stakeholder engagement (involving stakeholders in the systematic identification, analysis, planning, and implementation of actions for the digital innovation)</p>	<p>Missing cocreation (designers of digital innovation do not include medical, information technology, and business staff)</p> <p>Gap between technology developers/researchers and health care practice (involved parties have an opposing understanding or underestimate the time needed to complete relevant steps)</p> <p>Knowledge in silos (data and document systems of the digital innovation are not shared between all people involved)</p> <p>Missing common goal (goals of different stakeholders do not align)</p>
Innovation characteristics ^a	—	<p>Who pays the risk of innovation? (liability issues and uncertainties when digital health innovations are integrated into the treatment process)</p> <p>Too high pace of technology inventions (many technical improvements in a short time making it difficult to implement them owing to a steady flow of new technologies)</p> <p>Need for speed (rapidity/pressure of change)</p> <p>The “not invented here” dilemma (successful and effective digital innovations are not implemented when users or clinicians were not involved)</p>
Methodology ^b	—	Selection bias (certain types of professionals or clinicians are more interested in developing digital innovations; furthermore, certain users may be more eager to participate and use digital innovations)

^aAdditional category based on conclusions from the workshop.

^bCategory adopted from Kowatsch et al [3].

^cNot available.

Barriers rather than enablers were identified on the macro level, referring to legislation, regulation or finance guidelines, and the respective stakeholders. The characteristics of regulation, funding, reimbursement, and planning are currently mainly hindering a successful scale-up of digital health innovations. Liability issues or the generally missing innovation-friendliness in the health care system together with missing funding or reimbursement can be named as examples. Further, the experts shared their experience that the aim of most research is not a

successful implementation. All barriers and enablers on the macro level are presented in [Table 2](#).

The meso level, for example, the community around individual end users, on the contrary, seems to enable the scale-up of digital health innovations rather than hinder it. Even though using different infrastructure systems as regional infrastructure can be a barrier, culture and social support can help in raising awareness and building capacity and trust. All barriers and enablers on the meso level are presented in [Table 3](#).

Table 2. Enablers and barriers for scaling up digital health innovations on the macro level identified in the workshop and aligned with prior work.

Category ^a	Enabler	Barrier
Regulatory issues	A method for approval of market entry (guidelines and rules along which new digital innovations can be developed and introduced, eg, Digital Health Applications process from the German Federal Institute for Drugs and Medical Devices) [40] Legislative change (recent changes improving the development, implementation, and reimbursement of digital health innovations, eg, Digital Healthcare Act [Digitale-Versorgung-Gesetzes] passed on December 19, 2019, the “app on prescription” for patients was introduced into health care [sections §§ 33a and 139e of the Fifth Book of the German Social Code Book V])	Legal regulations (legislation that regulates development and market entry for digital health innovations) Liability issues (unclear who is responsible for the safety and possible claims that could arise in the future) High regulatory barriers (overhead to develop and distribute digital innovations is high) Health system is not innovation-friendly
Funding	— ^b	Missing funding (funding necessary for the development and accreditation of digital health innovations is often missing)
Reimbursement	—	Reimbursement is not guaranteed (unclear whether health care providers will get paid for using/prescribing the digital health innovation)
Planning	—	The aim of the research is not a successful implementation

^aAll categories were adopted from Kowatsch et al [3].

^bNot available.

Table 3. Enablers and barriers for scaling up digital health innovations on the meso level identified in the workshop and aligned with prior work.

Category ^a	Enabler	Barrier
Regional infrastructure	— ^b	Different infrastructure systems are used (eg, different database, digital health records, diagnosis, and treatment codes)
Culture	Organizational change (change from one state of affairs to another in the form of, eg, company’s structure, strategy, policies, procedures, technology, or culture) Capacity building (obtaining, improving, and retaining the skills knowledge, tools, equipment, and other resources needed to achieve organizations’ goals) Awareness raising (inform and educate individuals about a topic or issue with the intent to change their attitudes, behaviors and beliefs, eg, the potential of digital innovations for health) Prioritization of trustworthy digital health (professionals and users need to trust in the digital health innovations’ effectiveness and safety)	—
Social support	Trust building (increasing the users’ trust in the safety and effectiveness as well as the vision and mission of the digital innovation and the individuals or organizations involved)	—

^aAll categories were adopted from Kowatsch et al [3].

^bNot available.

On the micro level, the level of individual end users, the recommendation of digital health innovations by physicians (social support) can enable their scale-up. Other individual

characteristics of the end users, for example, lacking motivation or trust, their individual resources, or negative associations are rather hindering. The latter two mainly refer to the

physicians/professionals as end users as they could be affected by additional work or see digital health innovations as a threat, negatively influencing their scale-up. All barriers and enablers on the micro level are presented in [Table 4](#).

Technical aspects, particularly aspects regarding the innovation itself, form the largest group of enablers and barriers named by the experts. The categories related are information quality, usability, integration, interoperability, the business model, standards, and the innovation process itself. Although information quality (eg, open source) is seen as an enabler, usability is only seen as a barrier if not well-thought-out (eg, lack of ease of use, too high complexity). Integration and interoperability aspects can hinder the innovation if not sufficiently considered but can also support actively the

innovation's success. Providing incentives or added value and having the business model in mind already at an early stage are important aspects for successfully scaling up digital health interventions. If, on the contrary, no suitable business model or no sufficient value propositions exist, it can easily turn into a hindering factor. Standards are a category, which only supports the scale-up of digital health innovations if appropriately followed. Using existing infrastructure and aligning to existing standards can be an advantage of each digital health innovation. Flexibility and modularization in the innovation process can support the innovation's success too, while the pressure of change and the risk of innovations are hindering in some cases. All barriers and enablers related to technical aspects or regarding innovation itself are presented in [Table 5](#).

Table 4. Enablers and barriers for scaling up digital health innovations on the micro level identified in the workshop and aligned with prior work.

Category ^a	Enabler	Barrier
Social support	Recommendation of the digital health innovation by physicians	— ^b
Individual characteristics of end user	—	Lack of motivation to change/adapt (digital innovation fails to elicit behavior) Trust issues (users do not trust the digital innovation to be effective, safe, or useful) Additional work for medical staff (digital innovation does not facilitate but increase the workload of already busy health care professionals)
Negative associations	—	Physicians perceive digital health innovations as a threat/substitution (digital health innovations as a potential replacement or restriction of professional latitude)

^aAll categories were adopted from Kowatsch et al [3].

^bNot available.

Table 5. Enablers and barriers related to technical aspects or regarding innovation itself for scaling up digital health innovations identified in the workshop and aligned with prior work.

Category	Enabler	Barrier
Information quality ^a	<p>Open source (source code of the digital innovations is made available for possible use, modification, and redistribution)</p> <p>Continuous clinical validation of digital innovations</p> <p>Information disclosure (the degree to what sensitive information is properly protected by the digital innovation)</p> <p>Evidence-based intervention components (components such as techniques, methods, and means to change behaviors or outcomes are based on empirical evidence)</p> <p>Access to patient data, software, etc (digital innovations need data from patients, access to software to be developed and validated on a larger scale)</p>	— ^b
Usability of technology ^c	—	<p>Lack of ease of use (digital innovation is burdensome)</p> <p>Complexity is too high (digital innovation targets too many outcomes or behaviors or is poorly designed)</p> <p>No user-centered design</p>
Integration ^c	Integration in existing workflows (digital health innovations can be integrated into existing systems for prevention or treatment)	Integration issues (digital innovations cannot or are difficult to integrate into existing systems or workflows)
Interoperability ^c	<p>Complemented and extended health care service delivery and research (does not compete with or disrupt workflow)</p> <p>Early steps on interoperability (digital innovations can exchange and use information and data from other sources)</p>	<p>Incompatibility of existing processes and innovation (digital innovations do not solve a problem in the current health care setting and cannot be used in other settings)</p> <p>Closed systems/missing interoperability (digital innovations are limited or cannot exchange and use information and data from other sources)</p>
Business model ^d	<p>Appropriate incentives (momentary or other compensations for participation in programs or using digital innovations)</p> <p>Financially viable business model (the degree to which digital innovations can cover costs and potentially generate revenue)</p> <p>Business model in mind at an early stage (of developing the digital health innovation)</p> <p>Providing added value (the degree to which the digital innovation can address the needs or ease the pains of users)</p>	<p>No suitable business model for preventive interventions</p> <p>Missing value proposition for patients (digital innovation does not solve or facilitate a problem or need of the patients)</p>
Standards ^c	<p>Alignment to existing standards (digital health innovation was developed by taking existing standards or guidelines into consideration)</p> <p>Usage of existing infrastructure (digital innovations are designed to use existing resources or incorporate relevant health care professionals)</p> <p>Utilization of existing organizations (patient organizations or research institutions are consulted when the digital innovation is designed)</p>	—
Innovation process ^d	<p>Minimum viable product and small iterations (small but working prototypes to be evaluated in continuous evaluation by users, health care providers, and health care professionals alike)</p> <p>Adoption, iteration, refinement, and removal of elements that do not add value</p> <p>Modularization regarding upscaling (further modules that extend the digital health innovation are developed and released)</p> <p>Flexibility in the innovation process (adjustment to findings from research and the design process are integrated)</p> <p>User-centered design and evaluation at every stage</p>	Unclear/not defined process to innovate (iteration of different stages of the digital health innovation or digital health innovations, in general, is not planned)
Interdisciplinary cocreation ^d	Patient inclusion (patients are integrated into each step when designing and evaluating the digital health innovation)	Missing broad stakeholder engagement (focusing on only one group)

^aCategory adopted from DeLone and McLean [37].

^bNot available.

^cCategory adopted from Kowatsch et al [3].

^dAdditional category based on conclusions from the workshop.

Discussion

Workshop outcomes were used to compile a classified list of enablers and barriers. The workshop's aim was to match existing theoretical insights on enablers and barriers of digital health innovations with the practical experiences that the workshop participants brought to the activity. Participants offered insights from both research-based (empirical and applied) and real-life perspectives. After the workshop, the resulting brainstorming (brainwriting) results were classified into categories to reach a single consolidated list of enablers and barriers. The results represent the collective perspective of the group members who participated in the working session. The viewpoint represented is that of a set of people who have actively participated in different forms of research on the development and implementation of digital health interventions.

The categories of enablers and barriers have shown that the successful scale-up of digital health innovations is influenced by actors and aspects on different levels (micro, meso, macro, and technology/innovation level). Actors in each of these levels are perceived to influence the success of digital health innovations. These different levels are in line with the focus areas that were identified by Labrique et al [41] critical for scaling digital health initiatives: health care ecosystem (macro level), extrinsic ecosystem (meso level), intrinsic characteristics, human factors (micro level), and technical factors (technology/innovation). Our viewpoint contributed to further specify, in each level, enablers and barriers experienced by different stakeholders.

Prior work of some of the participants in the workshop [3] covers diverse categories of influencing factors on digital health innovations. Although most of these categories were mirrored by the results of the workshop, some were not named at all by the workshop participants, as is the example of the disease, social interaction, or expectations. Other categories such as standards or social support were only named as enablers even though they also represent barriers according to Kowatsch et al [3]. However, it is a noteworthy fact that some of the aspects pointed out by the workshop participants add a new contribution to our prior work, namely, the characteristics and the process of innovations, leadership, interdisciplinary cocreation, and the business model.

Leadership was referred to as both enabler and barrier (when considering the lack of it). It is an essential trigger for digital innovation and adoption [42]. In line with leadership, the culture for change, the need for common goals, prioritization, and planning were mentioned in the workshop as relevant for scale-up. These relate to the intimate connection between digital health innovation and the changes that it drives in delivering health care services [43]. Scaling up digital health must be driven by synergic interventions in health care services and workflows. This justifies the cocreation interdisciplinary category mentioned in the workshop, reinforcing the need of

the different stakeholders to be engaged and to be considered. Cresswell et al [44] suggest 10 key considerations for the case of health information technology, where including professional, administrative, and managerial teams to define the needs and build consensus around a strategic vision is key to successfully implement technology. A less siloed approach that motivates interdisciplinary cocreation is referred to in both our workshop findings and the literature [41].

Further, commonly mentioned in the workshop, at different levels, are the regulatory and trust issues that hamper adoption at a large scale. Indeed, policymakers and health care management need to be part of innovation processes, understand the needs for change, and provide guidelines that increase trust to take decisions on scaling up [42]. Business models were also mentioned in our workshop as a category that is relevant for scaling up. This relates to the need to develop a multistakeholder perspective on value delivery in the health care ecosystem [45]. Further research should develop guidelines that consider the different levels and categories that were discussed in this paper to support technology innovators and providers in planning the scale-up of digital health innovations.

It needs to be elaborated further if the discrepancy between the findings of our workshop and prior literature shows different perspectives in the perception of enablers and barriers for digital health innovations between research and practice. Moreover, future research should corroborate our findings with larger groups of experts to see how they hold or vary per medical subindustry. We suggest that the development of digital health apps for smartphones may be different from the development of hospital information systems. The findings reported are mainly limited by the size of the expert group involved in the workshop. It could have been too small or not representative enough since the group was composed mainly of researchers. Nevertheless, the experts involved combined extensive experience in (digital) health care.

Importance of COVID-19 to Scale-up

The workshop took place at the very onset of the pandemic breakout; days after it occurred, Malta restricted travel in and out of the island. As so, the reported viewpoint did not consider all the digital transformation that happened in the last year owing to the need for fast response to health care needs. Many identified barriers for adoption and scale-up were suddenly put aside owing to the urgent need to provide remote and safe care and monitoring. Leadership, regulatory issues, and reimbursement are categories of those barriers, which were also identified in Table 1. Governments incentivized telecare as a digital-first pathway, followed by a rapid change in the insurance companies that reimbursed teleconsultations. Many countries relaxed privacy and data protection regulations during the crisis under General Data Protection Regulation exceptions for public interest [46]. Cooperation and evidence of the benefits for the patients were enablers, identified in this viewpoint, for the fast scale-up and adoption of digital health. However, digital health

interventions played a marginal role owing to the inadequacy of protocols and lack of readiness for implementation [47], which may be mainly related to the categories of barriers identified in Table 1: innovation characteristics, planning, integration, and culture. Digital adoption by health care was held by those technologies that were already mature, commonly used and that could be integrated into existing workflows. One example was the massive adoption of videocall platforms and instant messaging apps that were provided for teleconsultations as urgent replacements for usual clinical consultations to respond to the population questions and concerns of the general public [48]. In addition to changes taking place in health care organizations, citizens themselves scaled up the adoption of smart devices that were already on the market [49,50]. Symptom-checking and contact-tracing apps were downloaded to millions of smartphones under the polemic of data privacy versus population safety [51]. This may be related to increased motivation and trust by the population, listed as enablers at a micro level in Table 1. The COVID-19 pandemic has led to a rapid scale-up of telehealth services. Two strong enablers included the urgent and immediate demand from health and care systems as well as the population at large and the readiness of global technology companies that were ready to adapt their technologies to respond to the needs of the context. In the crisis, these enablers demonstrated how one can, in reality, suspend barriers that have previously been identified as causes of delay in digital adoption [52]. Although some of the barriers identified in this viewpoint may have been reduced (eg, lack of motivation to change), others were simply ignored temporarily and remain to be tackled in the future (eg, legal regulations).

Conclusions

To summarize the findings and discussions from the workshop, 5 conclusions for the scale-up of digital health innovations can

be highlighted. First, digital health services can help to drive data quality, outreach to communities, and manage disease transmission/progression. Second, to reach these aims, a general cultural shift is needed when aiming to have digital services as viable instruments in health care along with classic pharmaceutical, surgical, or other therapeutic measures. Third, technological developments and interoperability appear to be enablers supporting digital health services rather than acting as hindrances. This latter finding is rather surprising since lack of interoperability has often been named as a barrier in prior work [53,54]. Indeed, the European Commission has long called for a more extensive focus on interoperability to facilitate the increased use of digital health technologies. Fourth, when scaling up digital health innovations, it is important to ensure the involvement of all stakeholders, including people from different professions and occupations, and especially patients and citizens. Only through a joint effort on the part of all stakeholders can digital health services succeed. Fifth, the innovation process itself also plays a crucial role, especially in relation to culture and leadership. The innovation process should be partitioned into different stages. Within each stage, further research should examine how to best fulfill the respective stage. Innovation processes should also be considered in reimbursement models for digital health innovations to ensure that new technologies such as digital pills have a chance to be tested in real-world settings. When working on all the 5 aspects of enablers and barriers to digital innovation in health and care, we believe that the scale-up of digital health services can be strongly supported. A clear view of these aspects would guide the application of the growing funding for health care information technology toward accelerating its impact in health care services.

Acknowledgments

The workshop was cofunded by the Swiss Lung Association and the Swiss health insurer CSS. The work for this paper was also cofunded by the European Social Funds and the Free State of Saxony (grant 100310385).

Conflicts of Interest

Author LAM is affiliated with Pear Therapeutics, Inc, HealthSim, LLC, and Square2 Systems, Inc. Conflicts of interest are extensively managed by her academic institution, Dartmouth College. GWT, JO, and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology and Economics at ETH Zurich and the Institute of Technology Management at the University of St. Gallen, which is funded in part by the Swiss health insurer CSS. TK is also cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. Neither CSS nor Pathmate Technologies was involved in the study design, methods, or results discussion of this paper.

References

1. Agarwal R, Gao G, DesRoches C, Jha AK. The digital transformation of healthcare: current status and the road ahead. *Information Systems Research* 2010 Dec;21(4):796-809. [doi: [10.1287/isre.1100.0327](https://doi.org/10.1287/isre.1100.0327)]
2. Kvedar JC, Fogel AL, Elenko E, Zohar D. Digital medicine's march on chronic disease. *Nat Biotechnol* 2016 Mar;34(3):239-246. [doi: [10.1038/nbt.3495](https://doi.org/10.1038/nbt.3495)] [Medline: [26963544](https://pubmed.ncbi.nlm.nih.gov/26963544/)]
3. Kowatsch T, Otto L, Harperink S, Cotti A, Schlieter H. A design and evaluation framework for digital health interventions. *IT-Information Technology*. (5-6) 2019;61:253-263. [doi: [10.1515/itit-2019-0019](https://doi.org/10.1515/itit-2019-0019)]
4. Draguet V. eHealth: digital health and care overview. European Commission. URL: https://ec.europa.eu/health/ehealth-digital-health-and-care/overview_en [accessed 2020-04-02]

5. Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015 Feb 26;372(9):793-795. [doi: [10.1056/nejmp1500523](https://doi.org/10.1056/nejmp1500523)]
6. Hekler E, Tiro J, Hunter C, Nebeker C. Precision health: the role of the social and behavioral sciences in advancing the vision. *Ann Behav Med* 2020 Nov 01;54(11):805-826 [FREE Full text] [doi: [10.1093/abm/kaa018](https://doi.org/10.1093/abm/kaa018)] [Medline: [32338719](https://pubmed.ncbi.nlm.nih.gov/32338719/)]
7. Carvalho JV, Rocha, van de Wetering R, Abreu A. A Maturity model for hospital information systems. *Journal of Business Research* 2019 Jan;94:388-399. [doi: [10.1016/j.jbusres.2017.12.012](https://doi.org/10.1016/j.jbusres.2017.12.012)]
8. Carvalho JV, Rocha, Abreu A. Maturity models of healthcare information systems and technologies: a literature review. *J Med Syst* 2016 Apr 15;40(6):1-10. [doi: [10.1007/s10916-016-0486-5](https://doi.org/10.1007/s10916-016-0486-5)]
9. Cowie MR, Blomster JI, Curtis LH, Duclaux S, Ford I, Fritz F, et al. Electronic health records to facilitate clinical research. *Clin Res Cardiol* 2016 Aug 24;106(1):1-9. [doi: [10.1007/s00392-016-1025-6](https://doi.org/10.1007/s00392-016-1025-6)]
10. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. *npj Digital Med* 2018 May 8;1(1):1-10. [doi: [10.1038/s41746-018-0029-1](https://doi.org/10.1038/s41746-018-0029-1)]
11. Miotto R, Li L, Kidd BA, Dudley JT. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci Rep* 2016 May 17;6(1):1-10. [doi: [10.1038/srep26094](https://doi.org/10.1038/srep26094)]
12. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. *J Telemed Telecare* 2016 Oct 16;24(1):4-12. [doi: [10.1177/1357633x16674087](https://doi.org/10.1177/1357633x16674087)]
13. 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](https://doi.org/10.1056/nejmp2003539)]
14. Sim I. Mobile devices and health. *N Engl J Med* 2019 Sep 05;381(10):956-968. [doi: [10.1056/nejmra1806949](https://doi.org/10.1056/nejmra1806949)]
15. Coravos A, Khozin S, Mandl KD. Developing and adopting safe and effective digital biomarkers to improve patient outcomes. *npj Digit. Med* 2019 Mar 11;2(1):1-5. [doi: [10.1038/s41746-019-0090-4](https://doi.org/10.1038/s41746-019-0090-4)]
16. Barata F, Tinschert P, Rassouli F, Steurer-Stey C, Fleisch E, Puhan MA, et al. Automatic recognition, segmentation, and sex assignment of nocturnal asthmatic coughs and cough epochs in smartphone audio recordings: observational field study. *J Med Internet Res* 2020 Jul 14;22(7):e18082 [FREE Full text] [doi: [10.2196/18082](https://doi.org/10.2196/18082)] [Medline: [32459641](https://pubmed.ncbi.nlm.nih.gov/32459641/)]
17. Ma T, Sharifi H, Chattopadhyay D. Virtual humans in health-related interventions: A meta-analysis. 2019 Presented at: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems; May 4-9; Glasgow, Scotland, UK. [doi: [10.1145/3290607.3312853](https://doi.org/10.1145/3290607.3312853)]
18. 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](https://doi.org/10.2196/15823)] [Medline: [32039810](https://pubmed.ncbi.nlm.nih.gov/32039810/)]
19. 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;3:1-3 [FREE Full text] [doi: [10.1038/s41746-019-0215-9](https://doi.org/10.1038/s41746-019-0215-9)] [Medline: [32025572](https://pubmed.ncbi.nlm.nih.gov/32025572/)]
20. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. *World Psychiatry* 2019 Feb;18(1):20-28 [FREE Full text] [doi: [10.1002/wps.20610](https://doi.org/10.1002/wps.20610)] [Medline: [30600624](https://pubmed.ncbi.nlm.nih.gov/30600624/)]
21. Vernmark K, Hesser H, Topooco N, Berger T, Riper H, Luuk L, et al. Working alliance as a predictor of change in depression during blended cognitive behaviour therapy. *Cogn Behav Ther* 2019 Jul;48(4):285-299. [doi: [10.1080/16506073.2018.1533577](https://doi.org/10.1080/16506073.2018.1533577)] [Medline: [30372653](https://pubmed.ncbi.nlm.nih.gov/30372653/)]
22. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAs) in mobile health: key components and design principles for ongoing health behavior support. *Ann Behav Med* 2018 May 18;52(6):446-462 [FREE Full text] [doi: [10.1007/s12160-016-9830-8](https://doi.org/10.1007/s12160-016-9830-8)] [Medline: [27663578](https://pubmed.ncbi.nlm.nih.gov/27663578/)]
23. Rogers E. *Diffusion of Innovations*, 5th edition. New York: Simon & Schuster; 2010.
24. Simmons R, Fajans P, Ghiron L. *Scaling-up Health Service Delivery: From Pilot Innovations to Policies and Programmes*. Switzerland: World Health Organization; 2007.
25. Azevedo S, Londral A. Digital innovation in outpatient healthcare delivery services: a common methodology to introduce IOT technologies in two use-cases. 2020 Presented at: Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Scale-IT-up; February 24-26; Valletta, Malta. [doi: [10.5220/0009394908210826](https://doi.org/10.5220/0009394908210826)]
26. Teepe G, Kowatsch T. Swiss francs seem to make insured move: comparing daily and monthly financial incentives of a scalable digital health intervention. 2020 Presented at: Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Scale-IT-up; February 24-26; Valletta, Malta p. 827-833. [doi: [10.5220/0009396308270833](https://doi.org/10.5220/0009396308270833)]
27. Ollier J, Kowatsch T. The doctor will see yourself now: review and discussion of a mass-market self-service technology for medical advice. 2020 Presented at: Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Scale-IT-up; February 24-26; Valletta, Malta p. 798-807. [doi: [10.5220/0009373607980807](https://doi.org/10.5220/0009373607980807)]
28. Kaczmarek S, Benedict M, Susky M. The nature of digital innovation and what can be learned for information systems management. 2020 Presented at: Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Scale-IT-up; February 24-26; Valletta, Malta. [doi: [10.5220/0009311607870797](https://doi.org/10.5220/0009311607870797)]

29. Otto L. IT-governance in integrated care: a risk-centered examination in Germany. 2020 Presented at: Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - Scale-IT-up; February 24-26; Valletta, Malta p. A-20. [doi: [10.5220/0009376708080817](https://doi.org/10.5220/0009376708080817)]
30. Marsch LA, Guarino H, Acosta M, Aponte-Melendez Y, Cleland C, Grabinski M, et al. Web-based behavioral treatment for substance use disorders as a partial replacement of standard methadone maintenance treatment. *J Subst Abuse Treat* 2014 Jan;46(1):43-51 [FREE Full text] [doi: [10.1016/j.jsat.2013.08.012](https://doi.org/10.1016/j.jsat.2013.08.012)] [Medline: [24060350](https://pubmed.ncbi.nlm.nih.gov/24060350/)]
31. Christensen DR, Landes RD, Jackson L, Marsch LA, Mancino MJ, Chopra MP, et al. Adding an Internet-delivered treatment to an efficacious treatment package for opioid dependence. *J Consult Clin Psychol* 2014 Dec;82(6):964-972 [FREE Full text] [doi: [10.1037/a0037496](https://doi.org/10.1037/a0037496)] [Medline: [25090043](https://pubmed.ncbi.nlm.nih.gov/25090043/)]
32. Campbell ANC, Nunes EV, Matthews AG, Stitzer M, Miele GM, Polsky D, et al. Internet-delivered treatment for substance abuse: a multisite randomized controlled trial. *Am J Psychiatry* 2014 Jun;171(6):683-690 [FREE Full text] [doi: [10.1176/appi.ajp.2014.13081055](https://doi.org/10.1176/appi.ajp.2014.13081055)] [Medline: [24700332](https://pubmed.ncbi.nlm.nih.gov/24700332/)]
33. Jensen L, Knarvik U, Pedersen C, Tangene W, Whitehouse D. Personalised blueprint for telemedicine deployment: validated and tested version. Momentum for Mainstreaming Telemedicine Deployment in Daily Practice (Grant Agreement No 297320). 2015. URL: <http://www.telemedicine-momentum.eu/resources-documents/> [accessed 2020-05-01]
34. Grooten L, Vrijhoef HJM, Calciolari S, Ortiz LGG, Janečková M, Minkman MMN, et al. Assessing the maturity of the healthcare system for integrated care: testing measurement properties of the SCIROCCO tool. *BMC Med Res Methodol* 2019 Mar 18;19(1):63 [FREE Full text] [doi: [10.1186/s12874-019-0704-1](https://doi.org/10.1186/s12874-019-0704-1)] [Medline: [30885141](https://pubmed.ncbi.nlm.nih.gov/30885141/)]
35. CDHI. 1st international workshop on best practices for scaling-up digital innovations in healthcare. Vimeo. 2020. URL: <https://vimeo.com/channels/1542084> [accessed 2022-02-09]
36. Heslin P. Better than brainstorming? Potential contextual boundary conditions to brainwriting for idea generation in organizations. *Journal of Occupational and Organizational Psychology* 2009;82(1):129-145. [doi: [10.1348/096317908x285642](https://doi.org/10.1348/096317908x285642)]
37. DeLone WH, McLean ER. The DeLone and McLean Model of Information Systems Success: A ten-year update. *Journal of Management Information Systems* 2014 Dec 23;19(4):9-30. [doi: [10.1080/07421222.2003.11045748](https://doi.org/10.1080/07421222.2003.11045748)]
38. Otto L. Implementing and scaling up telemedicine initiatives: beyond user-centeredness. 2019 Presented at: IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA); November 3-7; Abu Dhabi, United Arab Emirates. [doi: [10.1109/AICCSA47632.2019.9035232](https://doi.org/10.1109/AICCSA47632.2019.9035232)]
39. Using the quadruple helix approach to accelerate the transfer of research and innovation results to regional growth. Publication Office of the European Union. URL: <https://op.europa.eu/en/publication-detail/-/publication/6e54c161-36a9-11e6-a825-01aa75ed71a1> [accessed 2021-02-17]
40. Digital health applications (DiGA) database. Federal Institute for Drugs and Medical Devices. URL: https://www.bfarm.de/EN/Medical-devices/Tasks/Digital-Health-Applications/_node.html [accessed 2021-04-20]
41. Labrique AB, Wadhvani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. *Global Health* 2018 Nov 03;14(1):103 [FREE Full text] [doi: [10.1186/s12992-018-0424-z](https://doi.org/10.1186/s12992-018-0424-z)] [Medline: [30390686](https://pubmed.ncbi.nlm.nih.gov/30390686/)]
42. Schulman KA, Richman BD. Toward an effective innovation agenda. *N Engl J Med* 2019 Mar 07;380(10):900-901. [doi: [10.1056/nejmp1812460](https://doi.org/10.1056/nejmp1812460)]
43. Shaw J, Agarwal P, Desveaux L, Palma DC, Stamenova V, Jamieson T, et al. Beyond "implementation": digital health innovation and service design. *NPJ Digit Med* 2018;1:48 [FREE Full text] [doi: [10.1038/s41746-018-0059-8](https://doi.org/10.1038/s41746-018-0059-8)] [Medline: [31304327](https://pubmed.ncbi.nlm.nih.gov/31304327/)]
44. Cresswell KM, Bates DW, Sheikh A. Ten key considerations for the successful implementation and adoption of large-scale health information technology. *J Am Med Inform Assoc* 2013 Jun;20(e1):e9-e13 [FREE Full text] [doi: [10.1136/amiajnl-2013-001684](https://doi.org/10.1136/amiajnl-2013-001684)] [Medline: [23599226](https://pubmed.ncbi.nlm.nih.gov/23599226/)]
45. Schiavone F, Mancini D, Leone D, Lavorato D. Digital business models and ridesharing for value co-creation in healthcare: A multi-stakeholder ecosystem analysis. *Technological Forecasting and Social Change* 2021 May;166:120647. [doi: [10.1016/j.techfore.2021.120647](https://doi.org/10.1016/j.techfore.2021.120647)]
46. Peek N, Suján M, Scott P. Digital health and care in pandemic times: impact of COVID-19. *BMJ Health Care Inform* 2020 Jun 21;27(1):e100166. [doi: [10.1136/bmjhci-2020-100166](https://doi.org/10.1136/bmjhci-2020-100166)]
47. Berardi C, Antonini M, Genie MG, Cotugno G, Lanteri A, Melia A, et al. The COVID-19 pandemic in Italy: Policy and technology impact on health and non-health outcomes. *Health Policy Technol* 2020 Dec;9(4):454-487 [FREE Full text] [doi: [10.1016/j.hlpt.2020.08.019](https://doi.org/10.1016/j.hlpt.2020.08.019)] [Medline: [32895626](https://pubmed.ncbi.nlm.nih.gov/32895626/)]
48. Glaser JJ, Overhage M, Guptill J, Appleby C, Trigg D. What the pandemic means for health care's digital transformation. *Harvard Business Review*. URL: <https://hbsp.harvard.edu/product/H0624G-PDF-ENG> [accessed 2021-02-17]
49. Rabe B. New survey finds 70% of consumers improved home during COVID-19, more than half used smart devices. CISION PR Newswire. 2021. URL: <https://www.prnewswire.com/news-releases/new-survey-finds-70-of-consumers-improved-home-during-covid-19-more-than-half-used-smart-devices-301201817.html> [accessed 2021-03-05]

50. Paliwal A. Global wearable cardiac devices market to witness a tremendous growth due to COVID-19 pandemic - exclusive report 130 pages by Research Dive. Bloomberg. 2020. URL: <https://www.bloomberg.com/press-releases/2020-08-17/global-wearable-cardiac-devices-market-to-witness-a-tremendous-growth-due-to-covid-19-pandemic-exclusive-report-130-pages-by> [accessed 2021-03-03]
51. Jacob S, Lawarée J. The adoption of contact tracing applications of COVID-19 by European governments. Policy Design and Practice 2020 Nov 28:1-15. [doi: [10.1080/25741292.2020.1850404](https://doi.org/10.1080/25741292.2020.1850404)]
52. Ford D, Harvey JB, McElligott J, King K, Simpson KN, Valenta S, et al. Leveraging health system telehealth and informatics infrastructure to create a continuum of services for COVID-19 screening, testing, and treatment. J Am Med Inform Assoc 2020 Dec 09;27(12):1871-1877 [FREE Full text] [doi: [10.1093/jamia/ocaa157](https://doi.org/10.1093/jamia/ocaa157)] [Medline: [32602884](https://pubmed.ncbi.nlm.nih.gov/32602884/)]
53. Jang-Jaccard J, Nepal S, Alem L, Li J. Barriers for delivering telehealth in rural Australia: a review based on Australian trials and studies. Telemed J E Health 2014 May;20(5):496-504. [doi: [10.1089/tmj.2013.0189](https://doi.org/10.1089/tmj.2013.0189)] [Medline: [24801522](https://pubmed.ncbi.nlm.nih.gov/24801522/)]
54. iCOPS. URL: <http://www.icops.co.uk/products/cecops-tecs-products/> [accessed 2021-03-03]

Edited by JMIRPE Office, A Mavragani; submitted 25.09.20; peer-reviewed by L Wessel, E Sezgin; comments to author 24.12.20; revised version received 17.02.21; accepted 28.12.21; published 11.03.22.

Please cite as:

Schlieter H, Marsch LA, Whitehouse D, Otto L, Londral AR, Teepe GW, Benedict M, Ollier J, Ulmer T, Gasser N, Ultsch S, Wollschlaeger B, Kowatsch T

Scale-up of Digital Innovations in Health Care: Expert Commentary on Enablers and Barriers

J Med Internet Res 2022;24(3):e24582

URL: <https://www.jmir.org/2022/3/e24582>

doi:[10.2196/24582](https://doi.org/10.2196/24582)

PMID:[35275065](https://pubmed.ncbi.nlm.nih.gov/35275065/)

©Hannes Schlieter, Lisa A Marsch, Diane Whitehouse, Lena Otto, Ana Rita Londral, Gisbert Wilhelm Teepe, Martin Benedict, Joseph Ollier, Tom Ulmer, Nathalie Gasser, Sabine Ultsch, Bastian Wollschlaeger, Tobias Kowatsch. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 11.03.2022. 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.

Viewpoint

The Disclosure of Personally Identifiable Information in Studies of Neighborhood Contexts and Patient Outcomes

Andrew Graham Rundle^{1*}, DrPH; Michael David Miller Bader^{2,3*}, PhD; Stephen John Mooney^{4*}, PhD

¹Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY, United States

²Department of Sociology, College of Arts and Sciences, American University, Washington, DC, United States

³Department of Public Administration & Policy, School of Public Affairs, American University, Washington, DC, United States

⁴Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, United States

*all authors contributed equally

Corresponding Author:

Andrew Graham Rundle, DrPH

Department of Epidemiology

Mailman School of Public Health

Columbia University

722 west 168th street

New York, NY, 10032

United States

Phone: 1 2123057619

Email: agr3@cumc.columbia.edu

Abstract

Clinical epidemiology and patient-oriented health care research that incorporates neighborhood-level data is becoming increasingly common. A key step in conducting this research is converting patient address data to longitude and latitude data, a process known as geocoding. Several commonly used approaches to geocoding (eg, ggmap or the tidygeocoder R package) send patient addresses over the internet to web-based third-party geocoding services. Here, we describe how these approaches to geocoding disclose patients' personally identifiable information (PII) and how the subsequent publication of the research findings discloses the same patients' protected health information (PHI). We explain how these disclosures can occur and recommend strategies to maintain patient privacy when studying neighborhood effects on patient outcomes.

(*J Med Internet Res* 2022;24(3):e30619) doi:[10.2196/30619](https://doi.org/10.2196/30619)

KEYWORDS

geocode; patient privacy; ethical conduct of research; disclosure; privacy; security; identification; health information; strategy; outcome; neighborhood

Introduction

Background

Imagine if a clinical researcher were to disclose a list of patient addresses to a third party that was outside of their hospital or health system without a formal agreement with that third party to secure the data. Imagine they then publicly announced that they disclosed the addresses, that the addresses belonged to patients with a specific disease, and that those patients were being treated at a specific hospital. The researcher's institutional review board (IRB) and Health Insurance Portability and Accountability Act (HIPAA) compliance office would be outraged at these violations of patient privacy. Yet, this sequence of events can happen inadvertently when studying how neighborhood conditions such as access to medical facilities or

surrounding food environments affect clinical outcomes in certain patient populations.

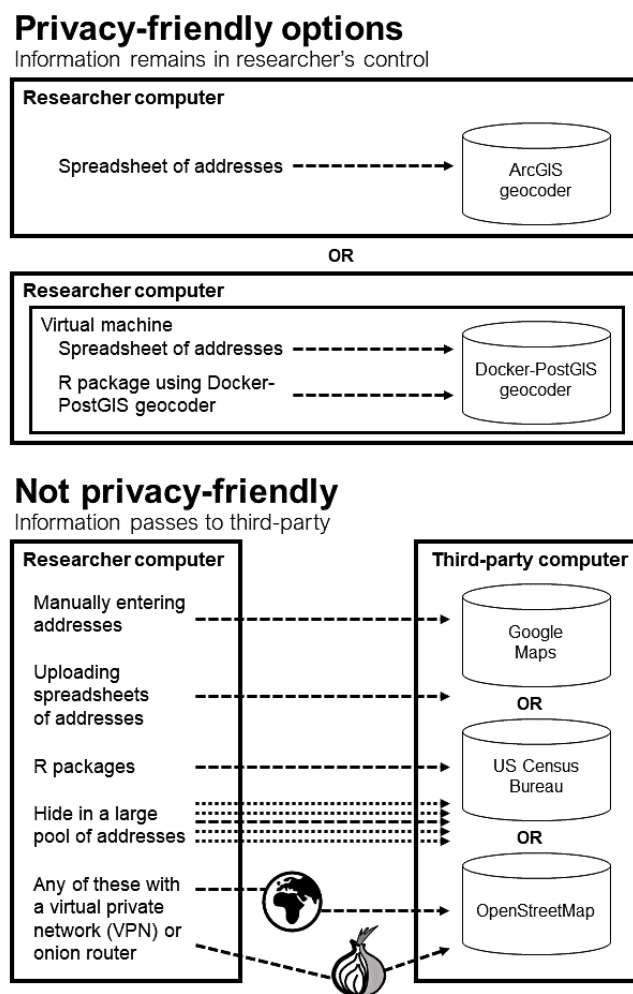
Clinical epidemiology and patient-oriented health care research incorporating neighborhood-level data is becoming progressively more common as the National Institutes of Health and professional and patient organizations are increasingly encouraging such research [1,2]. For instance, the American Cancer Society has classified multilevel research on neighborhood-level social determinants of cancer survivorship as a priority recommendation for research [2]. Here, we describe how inadvertent disclosures of personally identifiable information (PII) and protected health information (PHI) can occur when researchers study the effects of neighborhood contexts on clinical outcomes among patients and we describe ways to mitigate this risk. PII is information that can be used

to identify individuals; street addresses are classified as PII [3]. PHI includes information about a patient’s past or present health conditions, any treatment for those conditions, or any other provision of care. Under HIPAA, all geographical identifiers below the state level that were created, used, or disclosed during the provision of health care are also considered PHI [4]. Patient residential addresses used during treatment, or even addresses already documented in the medical or billing records, meet the criteria for geographic PHI, as does the name of the hospital treating a patient.

We describe the steps in the research process when PII and PHI can be disclosed when the study population is defined as patients with a specific disease or diseases, as is common in clinical

studies. We then explain why many protocols do not protect patient privacy and conclude by offering suggestions for a workflow that avoids the issues we identify. We use a fictional example throughout this article to avoid compounding potential PII and PHI disclosure issues that have occurred in previously published research. The hypothetical study being referenced examines whether residential neighborhood poverty rates are associated with the risk of death among patients with COVID-19 who were treated at the Columbia University Irving Medical Center. Though fictional, the example draws on published studies and studies that we have peer reviewed for clinical and public health journals. Figure 1 summarizes the disclosure risk associated with common approaches to geocoding patient addresses.

Figure 1. A summary of the personally identifiable information and protected health information disclosure risk for common approaches to geocoding patient addresses.



Privacy Pitfalls at Different Steps of Clinical Research

Step 1: Disclosure of Patient Addresses

The initial risk of disclosing PII and PHI occurs during a key step called geocoding. Geocoding is the process of converting a street address to longitude and latitude coordinates, which can then be linked to neighborhood-level data such as US census data. In our example of patients with COVID-19 who were treated at the Columbia University Irving Medical Center, the

patients’ addresses would be geocoded as the first step in estimating poverty rates in each patient’s residential neighborhood.

Several commonly used approaches to geocoding send addresses via the internet to third-party geocoding services. These approaches include manually entering addresses into websites such as Google Maps, uploading spreadsheets of addresses to geocoding websites, and using software packages such as the R (R Foundation for Statistical Computing) packages

tidygeocoder, gmap, and googleway; the Stata command geocode; and some options included with the SAS procedure GEOCODE. The tidygeocoder R package, for example, sends addresses to geocoding web services hosted by the US Census Bureau or by OpenStreetMap (or both), and gmap and googleway send addresses to the Google Maps geocoder [5,6]. The R package gmap provides easy access to the web-based Google Geocoding, Distance Matrix, and Directions application programming interface (API) services and is frequently used to geocode, map, and conduct spatial analyses of patient address data via Google Maps [6]. The proliferation of easy-to-use software procedures for geocoding and spatial analysis increases the risk of inadvertent disclosures of PII. A similar situation occurs when web-based geospatial tools such as Google Street View are used to measure patient neighborhood environments [7]. Revealing PII and PHI to third parties, such as those that host geocoding tools, violates typical human subjects research protocols approved by IRBs. The first 3 examples in the bottom panel of Figure 1 show how these options are unsafe as they pass data to servers outside of the control of the researcher, to an entity not covered by a Business Associate Agreement (BAA), a contractual obligation between 2 companies to safely handle HIPAA-protected data.

Step 2: Disclosure of the Identity of the Health Care System

An additional layer of disclosure occurs when using web-based geocoding services because these services can link home address data that were submitted to the service to the institution from which the data originated. When researchers send geocoding requests, geocoding services capture the IP address of the computer requesting the service along with the patient home address to be geocoded. As a result, the disclosed addresses are linked to a particular health care or research institution. Furthermore, the unique API keys used for many services create a linked history of all geocoding requests submitted by a researcher. Finally, data stored in the researcher's web browser cookies and search histories can also identify the researcher or clinician when they use some web-based geocoding services. Depending on which third-party service was used to geocode the residential addresses of the patients with COVID-19 in our example, the geocoding service provider would receive data on the identity of the institution originating the geocoding requests; they could also possibly receive data on the identity of the researcher and even information on the researcher's recent focus on COVID-19 research. Thus, the combination of patient address data, researcher or clinician identity, and the identity of the health institution allows the third-party geocoding service to make inferences about the submitted addresses.

Step 3: Disclosure of Patient Health Information

Publishing research results in articles allows third parties to further contextualize the disclosed addresses. Publication itself can therefore reveal PHI that can be linked to PII [3]. Publications describing the results of research about the effects of neighborhood contexts on outcomes in patient populations typically identify the geocoding service used, the number of addresses submitted to the service, the health system the patients belong to, and the health condition(s) used as criteria for

including patients in the study [8]. Publication of this study data provides sufficient information for the third party that geocoded the addresses to link patients' individual identifiers to their health conditions and health providers [3]. For example, researchers publishing the results of the hypothetical COVID-19 study would indicate how many patients were included in the study, the method used to geocode the sample (eg, tidygeocoder and OpenStreetMap), the number of addresses successfully geocoded, and the fact that the patients received treatment for COVID-19 at the Columbia University Irving Medical Center. The combination of server logs of addresses submitted for geocoding, the capture of the IP address of the institution that originated the geocoding requests, and the information in the published paper would be sufficient for the third-party geocoding service to identify the addresses as belonging to patients with COVID-19 who had been treated at the Columbia University Irving Medical Center.

Inadequate Strategies to Protect Patient Privacy

Some researchers recognize the risks associated with this type of study but use strategies that do not protect PII and PHI. One approach we have seen proposed is to submit patient addresses of interest to a geocoding service along with a pool of randomly selected addresses [9]. An issue with this approach is that due to referral patterns and the geographies of hospital catchment areas, patient addresses are likely to cluster and may not be sufficiently hidden by pools of randomly selected addresses. In theory, if the pool is large enough (eg, every address in the health care system's catchment or referral area), this approach to geocoding does not provide PII to the geocoding service provider. However, this approach has several flaws. First, the approach does not follow the National Institute of Standards and Technology recommendations to secure data by requiring an encryption key rather than relying on keeping the information itself hidden [10]. Second, identifying an address list long enough to effectively obscure the patient addresses typically requires more computational skill and resources than simply geocoding the addresses securely in the first place. In our example of outcomes among patients with COVID-19, this approach might require identifying and submitting to the geocoding service every address in the Columbia University Irving Medical Center patient catchment area to fully obscure the study's patient addresses.

Another possibility may be to use a virtual private network (VPN) that obscures the identity of the computer sending the addresses to the third-party geocoding service, thus concealing the fact that the addresses are being sent from a medical system. However, VPNs can be difficult to set up securely and many commercial VPNs do in fact reveal the location of the user's computer [11]. Furthermore, even with a secure VPN, default settings on web browsers reveal a user's location to the websites being visited. Thus, a researcher using a VPN to access a web-based geocoder still risks disclosing PII and PHI to the geocoder service provider. Moreover, even the successful implementation of a VPN only shifts the disclosure of the address information and researcher identity from the geocoding service provider to the VPN service provider. An onion router, such as Tor (The Onion Router), can add a layer of anonymity to ensure that the geocoding service provider cannot link the

address data to a researcher, clinician, or institution [12]. However, there have been cases of onion routers being compromised [13]. The complexity of these solutions makes full compliance difficult, especially since simpler solutions exist.

Appropriate Approaches to Geocoding Patient Data

For researchers to be compliant with HIPAA and typical IRB regulations, patient addresses should be geocoded using desktop tools, such as ArcGIS, that store address information on local secured machines. QGIS offers a free open-source alternative to ArcGIS, though default geocode options use OpenStreetMap, so care must be taken to set up desktop options. Alternatively, researchers can use a geocoder designed for server hosting within a virtual machine on their local computer. For example, the open-source PostGIS geocoder can be hosted within a Docker container and accessed using R (see [14] for an example). An alternative is for the hospital or health care system to negotiate a BAA with a third-party geocoding service [4]. Under HIPAA, a business associate provides services, including data analysis, that involve the use or disclosure of individually

identifiable health information to a covered entity [4]. Although licenses for ArcGIS can be expensive (note that the various third-party services and R packages we have described are free), without a BAA, the use of third-party services to geocode patient addresses involves the inappropriate disclosure of PII and PHI. The time and administrative cost of establishing a BAA may be equivalent to, or more expensive, than an ArcGIS license.

Conclusions

In summary, geocoding patient addresses using web-based services creates direct and large-scale disclosures of PII and PHI, and the proliferation of simple-to-use R packages for geocoding increases the risk of these inadvertent disclosures. Web-based geospatial tools are powerful and allow for efficient and rigorous research. However, researchers geocoding patient addresses must be aware of the risk of inadvertent disclosure of patient PII and PHI. Researchers should avoid web-based geocoders and use only secure desktop tools when geocoding patient information. Researchers may wish to disconnect computers from the internet while geocoding to ensure that no data are inadvertently passed to third parties.

Acknowledgments

AGR was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant P2CHD058486, awarded to the Columbia Population Research Center) and the National Institute of Diabetes and Digestive and Kidney Diseases (grant R01RDK127139). SM was supported by the National Library of Medicine (grant R00LM012868).

Conflicts of Interest

None declared.

References

1. Taplin SH, Anhang Price R, Edwards HM, Foster MK, Breslau ES, Chollette V, et al. Introduction: understanding and influencing multilevel factors across the cancer care continuum. *J Natl Cancer Inst Monogr* 2012 May;2012(44):2-10 [FREE Full text] [doi: [10.1093/jncimonographs/lgs008](https://doi.org/10.1093/jncimonographs/lgs008)] [Medline: [22623590](https://pubmed.ncbi.nlm.nih.gov/22623590/)]
2. Alcaraz KI, Wiedt TL, Daniels EC, Yabroff KR, Guerra CE, Wender RC. Understanding and addressing social determinants to advance cancer health equity in the United States: a blueprint for practice, research, and policy. *CA Cancer J Clin* 2020 Jan;70(1):31-46 [FREE Full text] [doi: [10.3322/caac.21586](https://doi.org/10.3322/caac.21586)] [Medline: [31661164](https://pubmed.ncbi.nlm.nih.gov/31661164/)]
3. McCallister E, Grance T, Scarfone K. Guide to protecting the confidentiality of personally identifiable information (PII): recommendations of the National Institute of Standards and Technology. National Institute of Standards and Technology. Gaithersburg, MD: US Department of Commerce; 2010. URL: https://tsapps.nist.gov/publication/get_pdf.cfm?pub_id=904990 [accessed 2021-07-14]
4. Office of Civil Rights. Summary of the HIPAA privacy rule. US Department of Health & Human Services. 2013. URL: <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html> [accessed 2021-07-01]
5. Cambon J, Hernangómez D, Belanger C, Possenriede D. tidygeocoder: geocoding made easy. The Comprehensive R Archive Network. URL: <https://cran.r-project.org/web/packages/tidygeocoder/index.html> [accessed 2021-02-01]
6. Kahle D, Wickham H. ggmap: spatial visualization with ggplot2. *The R Journal* 2013;5(1):144. [doi: [10.32614/rj-2013-014](https://doi.org/10.32614/rj-2013-014)]
7. Bader MDM, Mooney SJ, Rundle AG. Protecting personally identifiable information when using online geographic tools for public health research. *Am J Public Health* 2016 Feb;106(2):206-208. [doi: [10.2105/AJPH.2015.302951](https://doi.org/10.2105/AJPH.2015.302951)] [Medline: [26794375](https://pubmed.ncbi.nlm.nih.gov/26794375/)]
8. Jia P, Yu C, Remais JV, Stein A, Liu Y, Brownson RC, et al. Spatial lifecourse epidemiology reporting standards (ISLE-ReSt) statement. *Health Place* 2020 Jan;61:102243 [FREE Full text] [doi: [10.1016/j.healthplace.2019.102243](https://doi.org/10.1016/j.healthplace.2019.102243)] [Medline: [32329723](https://pubmed.ncbi.nlm.nih.gov/32329723/)]
9. Petitcolas F, Anderson R, Kuhn M. Information hiding—a survey. *Proc IEEE* 1999;87(7):1062-1078. [doi: [10.1109/5.771065](https://doi.org/10.1109/5.771065)]
10. Scarfone K, Jansen W, Tracey M. Guide to general server security: recommendations of the National Institute of Standards and Technology. National Institute of Standards and Technology. 2008. URL: <https://nvlpubs.nist.gov/nistpubs/legacy/sp/nistspecialpublication800-123.pdf> [accessed 2021-07-14]

11. Khan M, DeBlasio J, Voelker G, Snoeren A, Kanich C, Vallina-Rodriguez N. An empirical analysis of the commercial VPN ecosystem. In: Proceedings of the Internet Measurement Conference. 2018 Presented at: Internet Measurement Conference; October 31-November 2, 2018; Boston, MA. [doi: [10.1145/3278532.3278570](https://doi.org/10.1145/3278532.3278570)]
12. Reed M, Syverson P, Goldschlag D. Anonymous connections and onion routing. *IEEE J Select Areas Commun* 1998;16(4):482-494. [doi: [10.1109/49.668972](https://doi.org/10.1109/49.668972)]
13. Zetter K. Rogue nodes turn Tor anonymizer into eavesdropper's paradise. *Wired*. 2007. URL: <https://www.wired.com/2007/09/rogue-nodes-turn-tor-anonymizer-into-eavesdroppers-paradise/> [accessed 2021-07-14]
14. postgis-docker. GitHub. URL: <https://github.com/uwrit/postgis-docker> [accessed 2022-03-02]

Abbreviations

API: application programming interface
BAA: Business Associate Agreement
HIPAA: Health Insurance Portability and Accountability Act
IRB: Institutional Review Board
VPN: virtual private network
PHI: protected health information
PII: personally identifiable information
Tor: The Onion Router

Edited by A Mavragani; submitted 21.05.21; peer-reviewed by P James, K Kishore; comments to author 28.06.21; revised version received 12.08.21; accepted 22.12.21; published 17.03.22.

Please cite as:

Rundle AG, Bader MDM, Mooney SJ

The Disclosure of Personally Identifiable Information in Studies of Neighborhood Contexts and Patient Outcomes

J Med Internet Res 2022;24(3):e30619

URL: <https://www.jmir.org/2022/3/e30619>

doi: [10.2196/30619](https://doi.org/10.2196/30619)

PMID: [35103610](https://pubmed.ncbi.nlm.nih.gov/35103610/)

©Andrew Graham Rundle, Michael David Miller Bader, Stephen John Mooney. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 17.03.2022. 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.

Viewpoint

Gray Literature in Evaluating Effectiveness in Digital Health and Health and Welfare Technology: A Source Worth Considering

Sara Landerdahl Stridsberg¹, MA; Matt X Richardson², PhD; Ken Redekop³, PhD; Maria Ehn⁴, PhD; Sarah Wamala Andersson⁵, PhD

¹University Library, Mälardalen University, Västerås, Sweden

²School of Health and Welfare, Mälardalen University, Eskilstuna, Sweden

³Erasmus School of Health Policy and Management, Erasmus Universiteit Rotterdam, Rotterdam, Netherlands

⁴School of Innovation, Design, and Engineering, Mälardalen University, Västerås, Sweden

⁵School of Health and Welfare, Mälardalen University, Västerås, Sweden

Corresponding Author:

Sara Landerdahl Stridsberg, MA

University Library

Mälardalen University

Box 883

Västerås, S-721 23

Sweden

Phone: 46 21101557

Email: sara.landerdahl.stridsberg@mdu.se

Abstract

Background: The need to assess the effectiveness and value of interventions involving digital health and health and welfare technologies is becoming increasingly important due to the rapidly growing development of these technologies and their areas of application. Systematic reviews of scientific literature are a mainstay of such assessment, but publications outside the realm of traditional scientific bibliographic databases—known as gray literature—are often not included. This is a disadvantage, particularly apparent in the health and welfare technology (HWT) domain.

Objective: The aim of this article is to investigate the significance of gray literature in digital health and HWT when reviewing literature. As an example, the impact of including gray literature to the result of two systematic reviews in HWT is examined.

Methods: In this paper, we identify, discuss, and suggest methods for including gray literature sources when evaluating effectiveness and appropriateness for different review types related to HWT. The analysis also includes established sources, search strategies, documentation, and reporting of searches, as well as bias and credibility assessment. The differences in comparison to scientific bibliographic databases are elucidated. We describe the results, challenges, and benefits of including gray literature in 2 examples of systematic reviews of HWT.

Results: In the 2 systematic reviews described in this paper, most included studies came from context-specific gray literature sources. Gray literature contributed to the overall result of the reviews and corresponded well with the reviews' aims. The assessed risk of bias of the included studies derived from gray literature was similar to the included studies from other types of sources. However, because of less standardized publication formats, assessing and extracting data from gray literature studies were more time-consuming and compiling statistical results was not possible. The search process for gray literature required more time and the reproducibility of gray literature searches were less certain due to more unstable publication platforms.

Conclusions: Gray literature is particularly relevant for digital health and HWT but searches need to be conducted systematically and reported transparently. This way gray literature can broaden the range of studies, highlight context specificity, and decrease the publication bias of reviews of effectiveness of HWT. Thus, researchers conducting systematic reviews related to HWT should consider including gray literature based on a systematic approach.

(*J Med Internet Res* 2022;24(3):e29307) doi:[10.2196/29307](https://doi.org/10.2196/29307)

KEYWORDS

health and welfare technology; digital health; gray literature; information retrieval

Introduction

Reviewing Literature in Digital Health and Health and Welfare Technology

Connected or overlapping terms are often used to describe digital interventions in health and care services, including eHealth, mobile health (mHealth), digital health, telehealth, and telemedicine [1]. The term digital health has become an established umbrella term for these [2] and implies “the use of information and communication technologies to improve human health, health care services, and wellness for both individuals and populations” [3]. The term health and welfare technology (HWT) [4], now broadly used in the Nordic countries among others, adds more detail to the digital health concept. HWT is defined as “a technology-based intervention that aims at maintaining or promoting health, well-being, quality of life, and/or increasing efficiency in the service delivery system of welfare, social, and health care services, while improving working conditions of the staff” [5]. The combination of digital health and HWT thereby encompasses broad and burgeoning interdisciplinary fields of research that may not always prioritize established scientific literature databases when publicizing results.

The Cochrane Handbook for Systematic Reviews of Interventions, considered the gold standard in conducting such reviews, recommends that “searches for studies should be as extensive as possible in order to reduce the risk of publication bias and to identify as much relevant evidence as possible” [6]. Bibliographic databases mainly index studies published in peer-reviewed journals. Depending on the question and the scope of the review, however, the proportion of relevant studies not published in scholarly journals may vary greatly. Those not indexed in bibliographic databases are frequently referred to as gray literature.

What Is Gray Literature?

An often-cited definition of gray literature is “that which is produced on all levels of government, academics, business, and industry in print and electronic formats, but which is not controlled by commercial publishers (ie, where publishing is not the primary activity of the producing body)” [7]. Gray literature, according to this definition, is very diverse and could encompass self-published studies from research institutes, short conference abstracts, theses and dissertations, and ongoing research (trial registers), as well as government and committee reports, among others [8]. With today’s digital development and new media landscape, the related term gray data is sometimes used, including user-generated content such as blogs and social media [9].

Is It Worth the Trouble to Include Gray Literature in Your Search?

Methodological handbooks for literature reviews promote inclusion of gray literature to increase quality and depth of reviews [6,10,11]. This can be achieved by identifying ongoing studies with useful data as well as finding additional sources for evidence above and beyond what is normally found in commercially published material. Gray literature, therefore, has

the potential to reduce publication bias, which occurs when the published research is not representative of all the research that has been conducted [12].

Industry sponsors and technology developers may have different motives and timelines for work conducted in the field of digital health and HWT that may not coincide with traditional academic processes, scientific methods, or the conduct of rigorously controlled trials. In areas where the technology is developing fast, other study types than randomized controlled trials (RCT) may be preferred when evaluating interventions to allow for more rapid dissemination of results [13].

Even completed RCTs are not always published or retrievable. Al-Durra et al [14] found that 27% of all RCTs in digital health remained unpublished 5 years after completion. Industry-sponsored trials are less often published in scientific journals compared to nonsponsored or publicly funded trials, as found in a study of ClinicalTrials.gov [15]. Whether this is due to time constraints or unwillingness to widely disseminate findings, gray literature searches may reduce potential publication bias by identifying such unpublished, or ongoing, studies [16].

Moreover, rigorous and holistic assessments of evidence for interventions may not be found in bibliographic databases either. For example, publications from organizations that compile and assess evidence-based practice, such as the UK-based National Institute for Health and Care Excellence, and other health technology assessment (HTA) organizations will likely be missed if searches are restricted to bibliographic databases.

Another reason for including gray literature is that it is widely used. A recent study on nursing journals shows that gray literature accounts for about 10% of all citations [17], while another study showed that more than 10% of UK governmental publications in the field of health care are cited in the publication database Scopus [18]. Farrah and Mierzwinski-Urban [19] found that about 47% of the references on reports on novel nondrug health technologies were found in gray literature, and a review by Song et al [20] found that approximately half of all medical and health-related studies are not published in peer-reviewed journals. This can be particularly true in literature regarding new and emerging health technologies, but the reasons for not publishing studies can vary. In the Song et al [20] review, the main reason given for nonpublication was nonsubmission, where 85% of unpublished studies had not been submitted to journals due to lack of time or low priority. The next most common reason was that studies were incomplete or still ongoing. There may also be alternative incentives for publishing research in other sources than academic journals. In a study concerning the production of gray literature in the Australian public sector [21], the main reason organizations gave for publication output was to provide an evidence base to inform policy or practice, and the most effective channel for achieving this was felt to be publishing via the organization’s own website.

Gray literature can make a valuable contribution to the results in reviews of literature, although its potential relevance and contribution to the review’s results are often dependent upon the review type and its aim. Its strongest contribution may be in the conduct of scoping reviews, where it is a recommended

source to search along with scientific bibliographic databases. The aim of a scoping review is to provide an overview of a field of research, quite often with a focus on providing a general picture on the available literature for a topic [22,23] or understanding the context for an intervention [9]. Gray literature suits both these purposes well. As assessment of study quality or risk of bias is generally not performed [24], gray literature may also be more easily and reliably integrated alongside scientific publications in these types of reviews.

Including unpublished studies in meta-analyses, where studies in a systematic literature review are statistically analyzed, is generally considered to improve precision in the overall result. However, it is less clear if the inclusion of gray literature influences the effect size or statistical significance [25-27]. Another advantage of including gray literature in meta-analyses reported by Halfpenny et al [26] is that they found more data in trial registers than in the published study concerning the same trial; including gray literature therefore yielded a fuller and more complete result. Integration of some types of nontrial-based gray literature alongside published studies may be more challenging, however, as requirements on the rigor and format of statistical presentations can be highly variable in comparison.

Gray literature can also strengthen systematic literature reviews without meta-analyses, especially when the focus of the review is on evaluating how new interventions function in practice—an aspect particularly relevant for digital health and HWT. This may be particularly apparent in study areas where RCTs are not standard or common such as in interdisciplinary fields [28]—a description that also fits digital health and HWT well. Such areas may have a greater proportion of lower-quality research-based evidence and a higher context specificity that is important in implementation of interventions. This is where gray literature can provide a valuable supplement to other findings [29]. Examples of this include a study by Adams et al [9] regarding public health interventions, where most or all studies were found in gray literature, demonstrating its importance in investigations of effectiveness. In a systematic review of process and implementation of participatory ergonomics, Mahood et al [22] found that the several gray literature studies included alongside peer-reviewed literature resulted in a broader view of the evidence. Similarly, a systematic review by Cooper et al [30] of public health and environmental enhancement discovered that when comparing the results retrieved from searching in several bibliographic databases with a gray literature search (foremost via web searching) the latter contributed to a greater extent to the overall synthesis of the review. A result that might be related to the interdisciplinary nature of the subject for the review, sometimes not easily found when searching more subject-specific bibliographic databases.

There are disadvantages to using gray literature, however. One is the time and resources it takes to search for it. Even though the internet has made many types of gray literature more accessible via, for example, web pages and online repositories, gray literature can still be hard to find due to its comparative

deficits in metadata and indexing. The sustainability of gray literature digital domains is also a concern, with links and documents more prone to disappear than with bibliographic databases. Gray literature in many cases lacks time-saving functionality—for example, when exporting references to references management tools. It is also often less structured (eg, lacking abstracts), which makes any screening process less efficient [8,23,31].

Before undertaking a gray literature search, one should therefore consider whether the benefits outweigh the disadvantages. One example is the recommendation created by Jefferson et al [32] on whether to include the gray literature of regulatory data in Cochrane reviews concerning drugs or biologics. Key considerations include the burden of disease and number of people using or likely to use an intervention.

Because of the extra time required, Cochrane recommends limiting the use of gray literature when conducting rapid reviews and instead focusing mainly on study registers [33]. For the same reason, Cochrane recommends less frequent searches of gray literature than among publication databases when conducting and updating living reviews [34].

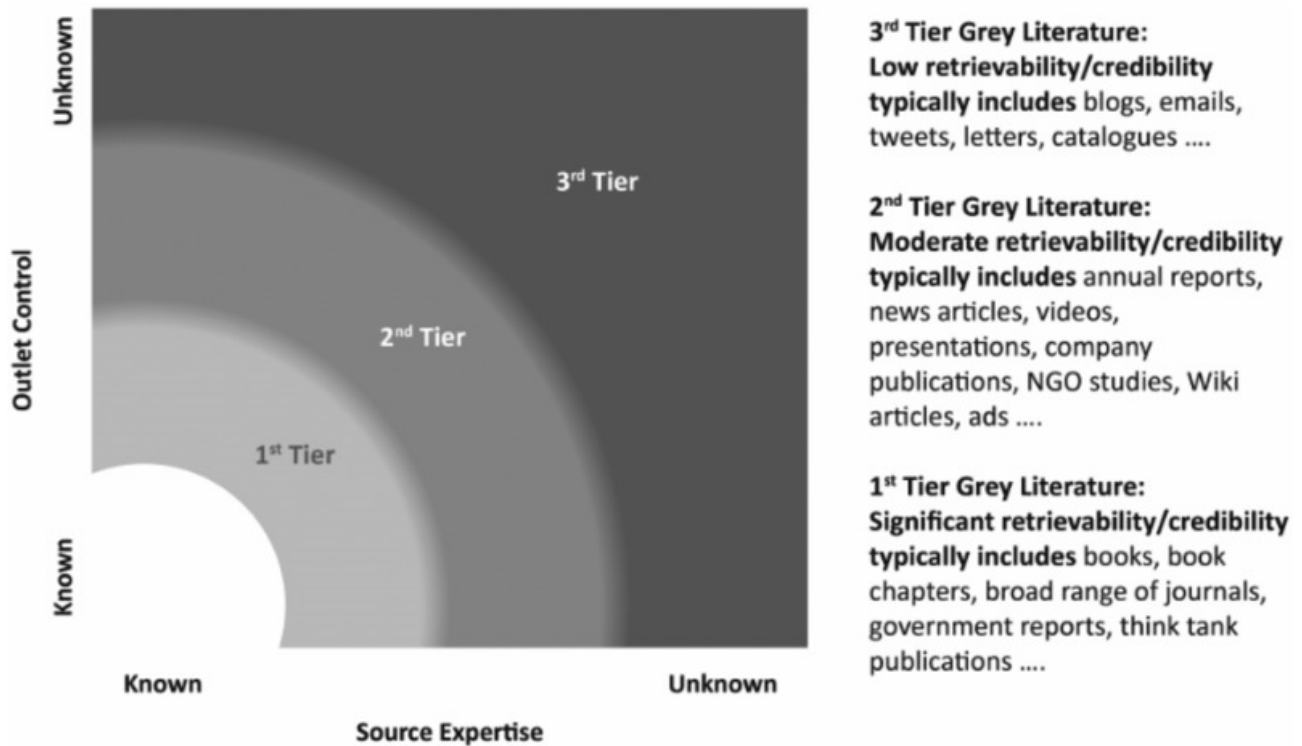
Is Gray Literature a Reliable and Credible Source to Use When Reviewing Literature?

Concerns about the quality of gray literature are often raised. Literature that in many cases has not been subjected to peer review and editorial examination is often viewed as lower quality and less credible. Some types of gray literature (eg, theses and governmental reports) do, however, receive a thorough quality check before being published [23]. In the previously mentioned study of Australian gray literature [21], two-thirds of the organizations stated that they had their work externally or internally reviewed prior to publication.

Like the evidence hierarchies within scientific literature, gray literature can be sorted according to the source's perceived usefulness. For example, Adams et al [16] described the connection between credibility, retrievability, and outlet control in gray literature by arranging it in 3 tiers (Figure 1). The first and most credible tier consists of government reports, books, and other highly traceable publication types, followed by a second tier that includes company publications and nongovernmental organization studies. The third and final tier includes publications with little or no outlet control, such as blogs and tweets, which would rarely be included in studies of effectiveness.

When assessing bias and study quality of scientific literature, tools like the Cochrane risk-of-bias 2.0 tool for randomized trials [35] and equivalents for other types of studies are well established. The same tools can certainly be used when assessing gray literature [22,36,37]. Other forms of assessment that focus more on the value of the information may be better suited to some kinds of gray literature [9]. One approach used in several reviews [38-40] is the AACODS checklist [41], which, in line with the list's acronym, evaluates the concepts of authority, accuracy, coverage, objectivity, date/time period, and significance of the content in a context specific to gray literature.

Figure 1. Tiers of gray literature credibility and retrievability in terms of source expertise and outlet control.



What Sources of Gray Literature Are Most Relevant for Digital Health and HWT?

Overview

An extensive list of sources to search, including gray literature, is available in the Technical Supplement to Chapter 4 in the Cochrane Handbook for Systematic Reviews of Interventions [42]. Campbell Collaboration provides similar guidance but focuses mainly on research fields related to education [43]. Both organizations advise researchers to consult a librarian or information specialist to identify relevant sources and formulate the search strategy [42,43].

As mentioned earlier, effectiveness studies in digital health and HWT may not always be carried out as RCTs. Non-RCT studies focusing on implementation and use, for example, may more often be found in practitioner-generated gray literature as opposed to academic gray literature [44]. While academic gray literature can be found in trial registers and theses, practitioner-generated gray literature is more likely to be found in organizational reports, government documents, and evaluations accessed via web searches. For an overview of sources of gray literature, see Table 1.

Table 1. Overview of sources of gray literature.

Type	Example
Practitioner-generated gray literature	
HTA ^a organizations	<ul style="list-style-type: none"> • Grey Matters, CADTH^b [45] • International HTA Database [46] • Evidence Search, NICE^c [47]
Web search engines	<ul style="list-style-type: none"> • Google [48] • Google Scholar [49] • Mednar [50] • DuckDuckGo [51]
Academic gray literature	
Dissertations, theses and academic papers	<ul style="list-style-type: none"> • OpenDOAR^d [52] • Open Access Theses and Dissertations [53] • ProQuest Dissertations and Theses Global [54] • DART Europe [55]
Study registers	<ul style="list-style-type: none"> • ClinicalTrials.gov [56] • WHO^e International Clinical Trials Registry Platform [57] • Finding Clinical Trials, Research Registers, and Research Results [58] • JMIR Research Protocols [59]
Conference papers	<p>Databases indexing conference papers:</p> <ul style="list-style-type: none"> • Zetoc [60] • Scopus [61] • Web of Science [62] • Embase [63] <p>Conference journals and websites:</p> <ul style="list-style-type: none"> • JMIR Iproceedings [64] • AMIA^f 2021 Virtual Annual Symposium [65]
Multidisciplinary gray literature databases	
Multiple sources (eg, institutional repositories, digital collections, and research reports)	<ul style="list-style-type: none"> • OpenGrey [66] • OAIster [67] • Bielefeld Academic Search Engine [68] • GreySource Index [69]

^aHTA: health technology assessment.

^bCADTH: Canadian Agency for Drugs and Technologies in Health.

^cNICE: National Institute for Health and Care Excellence.

^dDOAR: Directory of Open Access Repositories.

^eWHO: World Health Organization.

^fAMIA: American Medical Informatics Association.

Practitioner-Generated Gray Literature

Google [48] and Google Scholar [49] are often used when searching for gray literature on the web [70]. While Google is effective for retrieving information in many superficial searches, it does not cover the deep web content and therefore needs to be supplemented with searches of databases, specific websites, and organizational repositories [12,44]. Google also uses algorithms that provide or rank search results based on previous searches [31]. The incognito option available in some browsers can therefore be useful to avoid searches from becoming biased by personal preferences or habits [71]. Another solution to avoid algorithm bias is to use alternative search engines, such as Mednar [50], which has a medical focus, or DuckDuckGo [51],

a search engine that does not personalize based on previous searches.

HTA organizations evaluate digital health and HWT interventions and are thus a highly relevant source of gray literature. Grey Matters [45] from the Canadian Agency for Drugs and Technologies in Health lists HTA organizations from around the world as part of its checklist for searching gray literature.

Academic Gray Literature

Academic gray literature can be found in repositories for higher education, study registers, conference papers, and theses, among others. Study registers like ClinicalTrials.gov [56] and the World Health Organization (WHO) International Clinical Trials

Registry Platform [57] are mandatory in Cochrane reviews for interventions to avoid publication bias [6]. This may exclude digital HWT studies, however, as they may not meet the criteria for registration in clinical trial databases due to the nature of their intervention [72]. To make protocols from ongoing studies with different study designs more accessible, initiatives such as the journal JMIR Research Protocols [59] have been established and include protocols for ongoing studies and trials in digital health and related subjects.

Research conference papers often have the highest impact in fields with rapid knowledge development of new technologies and are thus highly valued sources [73]. In other areas, or when the available evidence is poor, conference abstracts can be worth considering if there is a lack of completed studies in a field [74]. Conference papers are indexed in some publication databases—for example, the multidisciplinary databases Scopus [61] and Web of Science [62] and the medical database Embase [63]. There are also dedicated databases for conference proceedings, for example Zetoc [60], as well as journals, for example the JMIR journal Iproceedings [64]. Many recurring conferences, such as the American Medical Informatics Association Annual Symposium [65], have websites for browsing and accessing papers. Some of these more established proceedings are starting to be indexed in databases such as PubMed, however, making them less gray and easier to find in scientific bibliographic database searches. The highly referenced Institute of Electrical and Electronics Engineers (IEEE) conference papers are also indexed in IEEE's own Xplore [75], which is considered a bibliographic database.

Academic papers can be found in national repositories for higher education such as OpenDOAR [52], a global directory of national academic repositories. International databases for dissertations and theses also exist such as Open Access Theses and Dissertations [53], ProQuest Dissertations and Theses Global [54], and Dart Europe [55], which has European content. Theses can be valuable and reliable sources of detailed knowledge on specific research topics as they disseminate results from lengthy periods of research and are subject to peer review [12]. They may also provide a more detailed background description and more extensive list of references than what is normally found in journal publications as they rarely have imposed word or page limits. Still, research that studied the impact of including theses on the result of systematic reviews from 3 Cochrane review groups found that in most reviews, their inclusion did not affect the review's results [76]. This could be explained by the fact that articles in several theses in major academic disciplines are also eventually published in scientific journals (ie, compilation theses).

Multidisciplinary Gray Literature Databases

There are databases dedicated to gray literature containing content from various sources—both academic and practitioner-based—and in many cases include functions for downloading references and conducting advanced searches, as in scientific bibliographic databases. A disadvantage may be that information on indexing and updating can be less transparent than for scientific bibliographic databases. There is also a greater risk that gray literature databases cease to be

updated as they often lack long-term funding and organization of commercial bibliographic databases. As with scientific bibliographic databases, searching several gray literature databases may be necessary to obtain an overlapping and more complete result [12]. The subject-specific databases for gray literature can be valuable sources for locating relevant literature but are dependent on the searcher's knowledge of the topic to identify them. Indexes for gray literature sources, such as GreySource Index [69], can therefore be valuable guides, serving as gateways to different sources of gray literature. GreySource Index is facilitated by GreyNet, an organization that provides a point of access to several repositories of gray literature.

Is It Possible to Be Systematic When Searching Gray Literature?

There are no generally established guidelines regarding methods and search strategies for gray literature [8,12,23]. Since the extent of gray literature is vast, making a search plan prior to the search helps estimate the time and resources needed to conduct the search.

The search plan should correspond to the aim of the search and identify the sources likely to yield the most value. Three key considerations when creating a search plan are as follows:

- Which types of studies/literature are of interest?
- Which methodological guidelines need to be followed?
- Which key databases, repositories, or websites are available on the topic?

Furthermore, time period and geographic coverage should be considered [12]. A systematic gray literature search, as with any other literature search, should use a clear population, intervention, comparison, and outcome format. Searches should also be as reproducible as possible, although gray literature searches often present challenges to this. Webpages and even repositories are less stable than scientific bibliographic databases and documents and links vanish or cease to exist more regularly.

How Should Searches for Gray Literature Be Documented?

Searches for gray literature should be documented and reported in a similar manner to searches in bibliographic databases. The aim of this documentation is to allow the search results to be reproduced by others. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [77] and extension PRISMA-S [71] provide guidance on how searches should be documented and reported when conducting a systematic review. The checklist Grey Matters [45] from the Canadian Agency for Drugs and Technologies in Health also includes elements of how to document and report a gray literature search.

According to PRISMA-S [71], gray literature searches should be briefly described in the methods section of the systematic review and complemented with more detailed information in supplementary material.

Example of Gray Literature Use in Two Systematic Reviews

Introduction

To illustrate how a search for gray literature can be conducted, we provide examples of methodology, benefits, and challenges taken from 2 systematic reviews we have published regarding the evidence for the use of welfare technologies: GPS alarms and digital nocturnal surveillance systems [36,37].

Aim of the Reviews

Both reviews aimed to find existing evidence for effects on health outcomes, welfare, and social care provision in elderly care for their respective technologies compared to standard care.

Methods

The inclusion criteria stated that original scientific and gray literature publications regarding the specific welfare technologies produced during the period 2005-2020 in Organisation for Economic Co-operation and Development (or equivalent) country settings were included. English, French, and Nordic languages were included in order to obtain a broader base of gray literature as many public sector reports were likely produced in non-English languages. Qualitative and proof-of-concept studies, technical validations, system descriptions, reviews, and editorials were excluded for both scientific and gray literature.

The scientific database searches were conducted first and involved abstract and full-text screening steps in accordance with PRISMA guidelines. Reference and citation searches (ie, backward and forward) were also conducted for relevant publications found. The databases searched for English-language publications were Academic Search Elite (EBSCOhost), APA PsycInfo (EBSCOhost), ASSIA (Applied Social Sciences Index and Abstracts, ProQuest), CINAHL Plus (EBSCOhost), Cochrane Library [78], IBSS (International Bibliography of the Social Sciences; ProQuest), IEEE Xplore [79], PubMed [80], Scopus [61], SocINDEX (EBSCOhost), Social Services Abstracts (ProQuest), Sociological Abstracts (ProQuest), Web of Science Core Collection [62].

The search terms used in the night surveillance review were (elderly OR “older adult*” OR “older person*” OR aged [MeSH only databases]) AND (nocturnal OR “night-time” OR “nighttime” OR “night time”) AND (surveillance OR camera* OR “video monitor*” OR “in-home monitor*” OR “home monitor*” OR “safety monitor*” OR “digital monitor*” OR telemonitor* OR “remote monitor*” OR “digital camera” OR “digital sensor*” OR “monitoring system*”). Subject terms were adapted to the controlled vocabulary of each database.

In the GPS review the following search string was used: (aged OR elder* OR “older adult*” OR “older person*” OR ageing OR aging OR senior*) AND (alarm*) AND (geofencing OR “global positioning system*” OR “positioning technolog*” OR “localization system*” OR “localization technolog*” OR “localisation system*” OR “localisation technolog*” OR “location tracking system*” OR “location tracking device*”).

OR “location tracking technolog*”). Also there, the search string was adapted to the subject headings of each database.

This was followed by initial searches in more established gray literature databases. These searches were conducted with the same search terms as for the scientific databases and in the English language as well. The main differences from the scientific database searches were that articles did not have to be peer reviewed, and the search strings (ie, combinations of search terms) had to be simplified in some cases as some sources’ search engines did not allow more complex strings with operator terms. In such cases, a greater number of searches consisting of fewer terms per search was required. The following resources were searched:

- Databases: Bielefeld Academic Search Engine (BASE) [68], OpenGrey [66], OAIster [67]
- HTA organization: International HTA database [46]
- Trial registers: ClinicalTrials.gov [56], WHO International Clinical Trials Registry Platform [57]
- Theses: Dart Europe [55], ProQuest Dissertations and Theses A&I [54]
- Web search engine: Google Scholar [49]

Additional searches were then performed in gray literature sources specific to the Nordic countries, where both GPS alarms and digital nocturnal surveillance systems have been implemented in social care. Due to the geographic focus, the inclusion criteria changed to studies from the Nordic countries and the searches were conducted in the Nordic languages. Repositories from higher education and websites from authorities responsible for elderly care and digitalization in health and welfare, municipalities, and research-based institutes were searched. A search was also made in Google Scholar with the respective search terms translated to the respective Nordic languages. The search string was adapted to the search functionality of the websites and search engines in a similar manner as the initial gray literature database searches.

Results

In the GPS alarm review, 56% (9/16) of studies came from gray literature sources (8 from the Nordic sources). In the digital nocturnal surveillance systems, 60% (3/5) of publications were from gray literature, all from Nordic sources. The same bias assessment tools (Cochrane’s Risk of Bias for randomized studies and the Risk Of Bias In Nonrandomized Studies of Interventions [ROBINS-I] tool for nonrandomized studies) were used for both scientific and gray literature studies. Bias was generally found to be moderate to critical, and the result was similar between the two types of sources.

Meta-analysis or consolidation of results was not possible due to lack of rigor in statistical reporting in the gray literature publications (eg, *P* values, confidence intervals, or standard errors not calculated or reported). Instead, a narrative summary with detailed reporting of included studies’ reported results was conducted. This provided adequate transparency for the reader when drawing conclusions about the overall effect of the technologies reviewed.

Discussion

Lessons Learned

The inclusion of gray literature in the two reviews brought both advantages and disadvantages, but the former clearly outweighed the latter. As most included studies came from gray literature, the volume of literature to base conclusions on was considerably larger—a systematic review of only two studies would have otherwise been the reality in one case. The included gray literature studies were also highly context-specific and had excellent adherence to the reviews' aims, perhaps even more so than the studies from the scientific database. It was, however, essentially impossible to consolidate results statistically in any meaningful way due to the lack of rigor in reporting statistics in the gray literature, which at least one journal reviewer deemed a disadvantage. It was also difficult to use the entire search strategy due to challenges with the search engines at the gray literature sources. Both the conduct of the gray literature searches and the subsequent import of found studies to the review software were far more time-consuming than in scientific publication databases; this was a far more manual process that could not use specialized applications that read the abundant metadata found in scientific databases to automate workflows. The time to review and extract data from gray literature studies was also far more time consuming due to less easily identifiable inclusion and exclusion criteria and less standardized publication formats. It is worth noting that, while the inclusion of Nordic

sources and specific language searches made the findings more context-specific, it may reduce the applicability of the results elsewhere. As well, reproducing the gray literature searches may be difficult if not impossible for other researchers due to the inability to interpret specific search results and the prospective fluctuation of available documents in the searched sources. Generally, reproduction of results on effectiveness is a challenge due to complexities in implementation of HWT in specific contexts.

Conclusion

Gray literature is worth considering when conducting many types of reviews related to HWT interventions as they can help reduce publication bias and include evidence for interventions that are not typically indexed in bibliographic databases. Including gray literature is particularly relevant for digital health and HWT, where studies tend to evaluate adoption and use using non-RCT study designs and for purposes other than primarily academic publishing. Caution needs to be taken when drawing conclusions from results in the gray literature due to a potential bias that may rise from less rigorous research design, methods, and data analyses. A risk of bias tool or checklist should be used, as with other literature. Finally, using a systematic approach when searching the gray literature is highly recommended. Planning the search and documenting it in a similar manner as when searching in bibliographic databases will save time and effort and make the search transparent and reproducible.

Acknowledgments

SWA formulated the idea for the manuscript. SLS wrote the paper. MXR, WKR, ME, and SWA all provided feedback and reviewed and edited the manuscript. This research was funded by the Mälardalen University research funding and the Sörmlandskontraktet regional funding partnership.

Conflicts of Interest

None declared.

References

1. Otto L, Harst L, Schlieter H, Wollschlaeger B, Richter P, Timpelditors. Towards a unified understanding of ehealth and related terms: proposal of a consolidated terminological basis. 2018 Presented at: HEALTHINF 11th International Conference on Health Informatics; 2018; Madeira. [doi: [10.5220/0006651005330539](https://doi.org/10.5220/0006651005330539)]
2. Kukafka R. Digital health consumers on the road to the future. *J Med Internet Res* 2019 Nov 21;21(11):e16359 [FREE Full text] [doi: [10.2196/16359](https://doi.org/10.2196/16359)] [Medline: [31750835](https://pubmed.ncbi.nlm.nih.gov/31750835/)]
3. Adjekum A, Blasimme A, Vayena E. Elements of trust in digital health systems: scoping review. *J Med Internet Res* 2018 Dec 13;20(12):e11254. [doi: [10.2196/11254](https://doi.org/10.2196/11254)] [Medline: [30545807](https://pubmed.ncbi.nlm.nih.gov/30545807/)]
4. Frennert S, Baudin K. The concept of welfare technology in Swedish municipal eldercare. *Disabil Rehabil* 2019 Sep 10:1-8. [doi: [10.1080/09638288.2019.1661035](https://doi.org/10.1080/09638288.2019.1661035)] [Medline: [31503509](https://pubmed.ncbi.nlm.nih.gov/31503509/)]
5. Andersson SW, Richardson MX, Cozza M, Lindén M, Redekop K. Addressing evidence in health and welfare technology interventions from different perspectives. *Health Policy Technol* 2021 Jun;10(2):100519. [doi: [10.1016/j.hlpt.2021.100519](https://doi.org/10.1016/j.hlpt.2021.100519)]
6. Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C, Metzendorf M. Chapter 4: Searching for and selecting studies. In: Higgins J, Chandler J, Cumpston M, Li T, editors. *Cochrane Handbook for Systematic Reviews of Interventions*. London: Cochrane; 2020.
7. Schopf J. Towards a Prague definition of grey literature. 2010 Presented at: Twelfth International Conference on Grey Literature: Transparency in Grey Literature. *Grey Tech Approaches to High Tech Issues*; 2010; Prague p. 11-26.
8. Paez A. Gray literature: an important resource in systematic reviews. *J Evid Based Med* 2017 Aug;10(3):233-240. [doi: [10.1111/jebm.12266](https://doi.org/10.1111/jebm.12266)] [Medline: [28857505](https://pubmed.ncbi.nlm.nih.gov/28857505/)]

9. Adams J, Hillier-Brown FC, Moore HJ, Lake AA, Araujo-Soares V, White M, et al. Searching and synthesising 'grey literature' and 'grey information' in public health: critical reflections on three case studies. *Syst Rev* 2016 Sep 29;5(1):164 [FREE Full text] [doi: [10.1186/s13643-016-0337-y](https://doi.org/10.1186/s13643-016-0337-y)] [Medline: [27686611](https://pubmed.ncbi.nlm.nih.gov/27686611/)]
10. Aromataris EM. *JBI Manual for Evidence Synthesis*. Adelaide: JBI Global; 2020.
11. Campbell C. *Campbell systematic reviews: policies and guidelines*. Campbell Collaboration. 2014. URL: <https://onlinelibrary.wiley.com/pb-assets/assets/18911803/Campbell%20Policies%20and%20Guidelines%20v4-1559660867160.pdf> [accessed 2022-02-22]
12. Bonato S. *Searching the Grey Literature: A Handbook for Searching Reports, Working Papers, and Other Unpublished Research*. Washington: Rowman & Littlefield Publishers; 2018.
13. Murray E, Hekler E, 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](https://doi.org/10.1016/j.amepre.2016.06.008)] [Medline: [27745684](https://pubmed.ncbi.nlm.nih.gov/27745684/)]
14. Al-Durra M, Nolan RP, Seto E, Cafazzo JA, Eysenbach G. Nonpublication rates and characteristics of registered randomized clinical trials in digital health: cross-sectional analysis. *J Med Internet Res* 2018 Dec 18;20(12):e11924 [FREE Full text] [doi: [10.2196/11924](https://doi.org/10.2196/11924)] [Medline: [30485832](https://pubmed.ncbi.nlm.nih.gov/30485832/)]
15. Chen CE, Harrington RA, Desai SA, Mahaffey KW, Turakhia MP. Characteristics of digital health studies registered in ClinicalTrials.gov. *JAMA Intern Med* 2019 Feb 25;838-840. [doi: [10.1001/jamainternmed.2018.7235](https://doi.org/10.1001/jamainternmed.2018.7235)] [Medline: [30801617](https://pubmed.ncbi.nlm.nih.gov/30801617/)]
16. Adams RJ, Smart P, Huff AS. Shades of grey: guidelines for working with the grey literature in systematic reviews for management and organizational studies. *Int J Manag Rev* 2016 Apr 19;19(4):432-454. [doi: [10.1111/ijmr.12102](https://doi.org/10.1111/ijmr.12102)]
17. Woods S, Phillips K, Dudash A. Grey literature citations in top nursing journals: a bibliometric study. *J Med Libr Assoc* 2020 Apr 01;108(2):262-269 [FREE Full text] [doi: [10.5195/jmla.2020.760](https://doi.org/10.5195/jmla.2020.760)] [Medline: [32256237](https://pubmed.ncbi.nlm.nih.gov/32256237/)]
18. Bickley MS, Kousha K, Thelwall M. Can the impact of grey literature be assessed? An investigation of UK government publications cited by articles and books. *Scientometrics* 2020 Jul 20;125(2):1425-1444. [doi: [10.1007/s11192-020-03628-w](https://doi.org/10.1007/s11192-020-03628-w)]
19. Farrah K, Mierzwinski-Urban M. Almost half of references in reports on new and emerging nondrug health technologies are grey literature. *J Med Libr Assoc* 2019 Jan 04;107(1):43-48 [FREE Full text] [doi: [10.5195/jmla.2019.539](https://doi.org/10.5195/jmla.2019.539)] [Medline: [30598647](https://pubmed.ncbi.nlm.nih.gov/30598647/)]
20. Song F, Loke Y, Hooper L. Why are medical and health-related studies not being published? A systematic review of reasons given by investigators. *PLoS One* 2014 Oct 15;9(10):e110418 [FREE Full text] [doi: [10.1371/journal.pone.0110418](https://doi.org/10.1371/journal.pone.0110418)] [Medline: [25335091](https://pubmed.ncbi.nlm.nih.gov/25335091/)]
21. Lawrence A. Influence seekers: the production of grey literature for policy and practice. *Inform Serv Use* 2018 Jan 08;37(4):389-403. [doi: [10.3233/isu-170857](https://doi.org/10.3233/isu-170857)]
22. Mahood Q, Van Eerd D, Irvin E. Searching for grey literature for systematic reviews: challenges and benefits. *Res Synth Methods* 2014 Sep 06;5(3):221-234. [doi: [10.1002/jrsm.1106](https://doi.org/10.1002/jrsm.1106)] [Medline: [26052848](https://pubmed.ncbi.nlm.nih.gov/26052848/)]
23. Gul S, Shah TA, Ahmad S, Gulzar F, Shabir T. Is grey literature really grey or a hidden glory to showcase the sleeping beauty. *Collection Curation* 2020 Feb 12;40(3):100-111. [doi: [10.1108/cc-10-2019-0036](https://doi.org/10.1108/cc-10-2019-0036)]
24. Peters MG, McInerney P, Munn Z, Tricco A, Khalil H. Chapter 11: Scoping reviews. *JBI Manual for Evidence Synthesis*. 2020. URL: <https://synthesismanual.jbi.global> [accessed 2022-02-22]
25. Baudard M, Yavchitz A, Ravaud P, Perrodeau E, Boutron I. Impact of searching clinical trial registries in systematic reviews of pharmaceutical treatments: methodological systematic review and reanalysis of meta-analyses. *BMJ* 2017 Feb 17;356:j448 [FREE Full text] [doi: [10.1136/bmj.j448](https://doi.org/10.1136/bmj.j448)] [Medline: [28213479](https://pubmed.ncbi.nlm.nih.gov/28213479/)]
26. Halfpenny NJA, Quigley JM, Thompson JC, Scott DA. Value and usability of unpublished data sources for systematic reviews and network meta-analyses. *Evid Based Med* 2016 Dec 29;21(6):208-213. [doi: [10.1136/ebmed-2016-110494](https://doi.org/10.1136/ebmed-2016-110494)] [Medline: [27686328](https://pubmed.ncbi.nlm.nih.gov/27686328/)]
27. Schmucker CM, Blümle A, Schell LK, Schwarzer G, Oeller P, Cabrera L, OPEN consortium. Systematic review finds that study data not published in full text articles have unclear impact on meta-analyses results in medical research. *PLoS One* 2017 Apr 25;12(4):e0176210 [FREE Full text] [doi: [10.1371/journal.pone.0176210](https://doi.org/10.1371/journal.pone.0176210)] [Medline: [28441452](https://pubmed.ncbi.nlm.nih.gov/28441452/)]
28. Cooper C, Booth A, Britten N, Garside R. A comparison of results of empirical studies of supplementary search techniques and recommendations in review methodology handbooks: a methodological review. *Syst Rev* 2017 Nov 28;6(1):234 [FREE Full text] [doi: [10.1186/s13643-017-0625-1](https://doi.org/10.1186/s13643-017-0625-1)] [Medline: [29179733](https://pubmed.ncbi.nlm.nih.gov/29179733/)]
29. Benzie KM, Premji S, Hayden KA, Serrett K. State-of-the-evidence reviews: advantages and challenges of including grey literature. *Worldviews Evid Based Nurs* 2006 Jun;3(2):55-61. [doi: [10.1111/j.1741-6787.2006.00051.x](https://doi.org/10.1111/j.1741-6787.2006.00051.x)] [Medline: [17040510](https://pubmed.ncbi.nlm.nih.gov/17040510/)]
30. Cooper C, Lovell R, Husk K, Booth A, Garside R. Supplementary search methods were more effective and offered better value than bibliographic database searching: a case study from public health and environmental enhancement. *Res Synth Methods* 2018 Jun 26;9(2):195-223. [doi: [10.1002/jrsm.1286](https://doi.org/10.1002/jrsm.1286)] [Medline: [29193834](https://pubmed.ncbi.nlm.nih.gov/29193834/)]
31. Godin K, Stapleton J, Kirkpatrick SI, Hanning RM, Leatherdale ST. Applying systematic review search methods to the grey literature: a case study examining guidelines for school-based breakfast programs in Canada. *Syst Rev* 2015 Oct 22;4(1):138 [FREE Full text] [doi: [10.1186/s13643-015-0125-0](https://doi.org/10.1186/s13643-015-0125-0)] [Medline: [26494010](https://pubmed.ncbi.nlm.nih.gov/26494010/)]

32. Jefferson T, Doshi P, Boutron I, Golder S, Heneghan C, Hodkinson A, et al. When to include clinical study reports and regulatory documents in systematic reviews. *BMJ Evid Based Med* 2018 Dec;23(6):210-217. [doi: [10.1136/bmjebm-2018-110963](https://doi.org/10.1136/bmjebm-2018-110963)] [Medline: [30309870](https://pubmed.ncbi.nlm.nih.gov/30309870/)]
33. Garritty C, Gartlehner G, Nussbaumer-Streit B, King VJ, Hamel C, Kamel C, et al. Cochrane Rapid Reviews Methods Group offers evidence-informed guidance to conduct rapid reviews. *J Clin Epidemiol* 2021 Feb;130:13-22 [FREE Full text] [doi: [10.1016/j.jclinepi.2020.10.007](https://doi.org/10.1016/j.jclinepi.2020.10.007)] [Medline: [33068715](https://pubmed.ncbi.nlm.nih.gov/33068715/)]
34. Brooker J, Synnot A, McDonald S, Elliott J, Turner T. Guidance for the production and publication of Cochrane living systematic reviews: Cochrane Reviews in living mode. 2019. URL: https://community.cochrane.org/sites/default/files/uploads/inline-files/Transform/201912_LSR_Revised_Guidance.pdf [accessed 2022-02-22]
35. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ* 2019 Aug 28;366:l4898. [doi: [10.1136/bmj.l4898](https://doi.org/10.1136/bmj.l4898)] [Medline: [31462531](https://pubmed.ncbi.nlm.nih.gov/31462531/)]
36. Ehn M, Richardson M, Landerdahl Stridsberg S, Redekop K, Wamala-Andersson S. Mobile safety alarms based on global positioning system technology in elderly care: a systematic review of evidence based on a general evidence framework for digital health technologies. *J Med Internet Res* 2021 Oct 11;23(10):e27267 [FREE Full text] [doi: [10.2196/27267](https://doi.org/10.2196/27267)] [Medline: [34633291](https://pubmed.ncbi.nlm.nih.gov/34633291/)]
37. Richardson MX, Ehn M, Stridsberg SL, Redekop K, Wamala-Andersson S. Nocturnal digital surveillance in aged populations and its effects on health, welfare and social care provision: a systematic review. *BMC Health Serv Res* 2021 Jun 30;21(1):622 [FREE Full text] [doi: [10.1186/s12913-021-06624-9](https://doi.org/10.1186/s12913-021-06624-9)] [Medline: [34187472](https://pubmed.ncbi.nlm.nih.gov/34187472/)]
38. Bandyopadhyay S, Baticulon RE, Kadhum M, Alser M, Ojuka DK, Badereddin Y, et al. Infection and mortality of healthcare workers worldwide from COVID-19: a systematic review. *BMJ Glob Health* 2020 Dec 04;5(12):e003097 [FREE Full text] [doi: [10.1136/bmjgh-2020-003097](https://doi.org/10.1136/bmjgh-2020-003097)] [Medline: [33277297](https://pubmed.ncbi.nlm.nih.gov/33277297/)]
39. Toppenberg-Pejcic D, Noyes J, Allen T, Alexander N, Vanderford M, Gamhewage G. Emergency risk communication: lessons learned from a rapid review of recent gray literature on ebola, zika, and yellow fever. *Health Commun* 2019 Apr 20;34(4):437-455. [doi: [10.1080/10410236.2017.1405488](https://doi.org/10.1080/10410236.2017.1405488)] [Medline: [29558199](https://pubmed.ncbi.nlm.nih.gov/29558199/)]
40. Johnson GA, Vindrola-Padros C. Rapid qualitative research methods during complex health emergencies: a systematic review of the literature. *Soc Sci Med* 2017 Sep;189:63-75. [doi: [10.1016/j.socscimed.2017.07.029](https://doi.org/10.1016/j.socscimed.2017.07.029)] [Medline: [28787628](https://pubmed.ncbi.nlm.nih.gov/28787628/)]
41. Tyndall J. AACODS checklist. 2010. URL: https://dspace.flinders.edu.au/xmlui/bitstream/handle/2328/3326/AACODS_Checklist.pdf [accessed 2022-03-18]
42. Lefebvre C, Glanville J, Briscoe S, Littlewood A, Marshall C. Technical Supplement to Chapter 4: Searching for and selecting studies. In: *Cochrane Handbook for Systematic Reviews of Interventions*. London: Cochrane Collaboration; 2020.
43. Kugley S, Wade A, Thomas J, Mahood Q, Jørgensen A, Hammerstrøm K, et al. Searching for studies: a guide to information retrieval for Campbell systematic reviews. *Campbell Systematic Reviews* 2017 Feb 13;13(1):1-73. [doi: [10.4073/cmg.2016.1](https://doi.org/10.4073/cmg.2016.1)]
44. Haddaway N, Bayliss H. Shades of grey: two forms of grey literature important for reviews in conservation. *Biological Conserv* 2015 Nov;191:827-829 [FREE Full text] [doi: [10.1016/j.biocon.2015.08.018](https://doi.org/10.1016/j.biocon.2015.08.018)]
45. Grey matters. Canadian Agency for Drugs and Technologies in Health. URL: <https://www.cadth.ca/grey-matters-practical-tool-searching-health-related-grey-literature> [accessed 2021-03-29]
46. International HTA Database. URL: <https://database.inahta.org> [accessed 2022-02-22]
47. Evidence Search, NICE. URL: <https://www.evidence.nhs.uk/> [accessed 2022-02-22]
48. Google. URL: <https://www.google.com/> [accessed 2021-03-29]
49. Google Scholar. URL: <https://scholar.google.com/> [accessed 2021-03-29]
50. Mednar. URL: <https://mednar.com/> [accessed 2021-03-29]
51. DuckDuckGo. URL: <https://duckduckgo.com/> [accessed 2021-03-29]
52. OpenDOAR. URL: <https://v2.sherpa.ac.uk/opensoar/> [accessed 2021-03-29]
53. Open Access Theses and Dissertations. URL: <https://oatd.org/> [accessed 2021-03-29]
54. ProQuest dissertations and theses A&I. URL: <https://about.proquest.com/products-services/pqdtglobal.html> [accessed 2022-02-22]
55. Dart Europe. URL: <https://www.dart-europe.org/basic-search.php> [accessed 2021-03-29]
56. ClinicalTrials.gov. URL: <https://clinicaltrials.gov> [accessed 2021-03-29]
57. WHO International Clinical Trials Registry Platform World Health Organisation. URL: <http://apps.who.int/trialsearch> [accessed 2021-03-29]
58. Finding Clinical Trials, Research Registers, and Research Results. URL: <https://sites.google.com/a/york.ac.uk/yhctrialsregisters/> [accessed 2022-02-22]
59. JMIR Research Protocols. URL: <https://www.researchprotocols.org> [accessed 2021-03-29]
60. Zetoc. URL: <http://zetoc.jisc.ac.uk> [accessed 2021-03-29]
61. Scopus. URL: <https://www.scopus.com> [accessed 2021-03-29]
62. Web of Science. URL: <https://www.webofknowledge.com> [accessed 2021-03-29]
63. Embase. URL: <https://www.embase.com/> [accessed 2022-02-22]
64. JMIR Iproceedings. URL: <https://www.iproc.org/> [accessed 2021-03-29]
65. AMIA 2021 Annual Symposium. URL: <https://www.amia.org/amia2021> [accessed 2021-03-29]

66. OpenGrey. URL: <http://www.opengrey.eu> [accessed 2022-02-22]
67. OAIster. URL: <https://www.oclc.org/en/oaister.html> [accessed 2022-02-22]
68. Bielefeld Academic Search Engine. URL: <https://www.base-search.net> [accessed 2022-02-22]
69. GreySource Index. URL: <http://greynet.org/greysourceindex.html> [accessed 2021-03-29]
70. Briscoe S, Nunns M, Shaw L. How do Cochrane authors conduct web searching to identify studies? Findings from a cross-sectional sample of Cochrane Reviews. *Health Info Libr J* 2020 Dec 08;37(4):293-318. [doi: [10.1111/hir.12313](https://doi.org/10.1111/hir.12313)] [Medline: [32511888](https://pubmed.ncbi.nlm.nih.gov/32511888/)]
71. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, PRISMA-S Group. PRISMA-S: an extension to the PRISMA Statement for Reporting Literature Searches in Systematic Reviews. *Syst Rev* 2021 Jan 26;10(1):39 [FREE Full text] [doi: [10.1186/s13643-020-01542-z](https://doi.org/10.1186/s13643-020-01542-z)] [Medline: [33499930](https://pubmed.ncbi.nlm.nih.gov/33499930/)]
72. Eysenbach G. Tackling publication bias and selective reporting in health informatics research: register your eHealth trials in the International eHealth Studies Registry. *J Med Internet Res* 2004 Sep 30;6(3):e35 [FREE Full text] [doi: [10.2196/jmir.6.3.e35](https://doi.org/10.2196/jmir.6.3.e35)] [Medline: [15471761](https://pubmed.ncbi.nlm.nih.gov/15471761/)]
73. Freyne J, Coyle L, Smyth B, Cunningham P. Relative status of journal and conference publications in computer science. *Commun ACM* 2010 Nov 01;53(11):124-132. [doi: [10.1145/1839676.1839701](https://doi.org/10.1145/1839676.1839701)]
74. Scherer RW, Saldanha IJ. How should systematic reviewers handle conference abstracts? A view from the trenches. *Syst Rev* 2019 Nov 07;8(1):264 [FREE Full text] [doi: [10.1186/s13643-019-1188-0](https://doi.org/10.1186/s13643-019-1188-0)] [Medline: [31699124](https://pubmed.ncbi.nlm.nih.gov/31699124/)]
75. IEEE Xplore. URL: <https://ieeexplore.ieee.org/Xplore/home.jsp> [accessed 2022-02-22]
76. Hartling L, Featherstone R, Nuspl M, Shave K, Dryden DM, Vandermeer B. Grey literature in systematic reviews: a cross-sectional study of the contribution of non-English reports, unpublished studies and dissertations to the results of meta-analyses in child-relevant reviews. *BMC Med Res Methodol* 2017 Apr 19;17(1):64 [FREE Full text] [doi: [10.1186/s12874-017-0347-z](https://doi.org/10.1186/s12874-017-0347-z)] [Medline: [28420349](https://pubmed.ncbi.nlm.nih.gov/28420349/)]
77. 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](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
78. Cochrane Library. URL: <https://www.cochranelibrary.com/> [accessed 2022-02-22]
79. IEEE Xplore. URL: <https://ieeexplore.ieee.org> [accessed 2022-02-22]
80. PubMed. URL: <https://pubmed.ncbi.nlm.nih.gov> [accessed 2022-02-22]

Abbreviations

AACODS: authority, accuracy, coverage, objectivity, date/time period, and significance

ASSIA: Applied Social Sciences Index and Abstracts

HTA: health technology assessment

HWT: health and welfare technology

IBSS: International Bibliography of the Social Sciences

IEEE: Institute of Electrical and Electronics Engineers

mHealth: mobile health

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

RCT: randomized controlled trial

ROBINS-I: Risk Of Bias In Nonrandomized Studies of Interventions

WHO: World Health Organization

Edited by A Mavragani; submitted 01.04.21; peer-reviewed by K Baudin, D Wu; comments to author 06.06.21; revised version received 08.07.21; accepted 02.12.21; published 23.03.22.

Please cite as:

Landerdahl Stridsberg S, Richardson MX, Redekop K, Ehn M, Wamala Andersson S

Gray Literature in Evaluating Effectiveness in Digital Health and Health and Welfare Technology: A Source Worth Considering

J Med Internet Res 2022;24(3):e29307

URL: <https://www.jmir.org/2022/3/e29307>

doi: [10.2196/29307](https://doi.org/10.2196/29307)

PMID: [35319479](https://pubmed.ncbi.nlm.nih.gov/35319479/)

©Sara Landerdahl Stridsberg, Matt X Richardson, Ken Redekop, Maria Ehn, Sarah Wamala Andersson. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 23.03.2022. 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.

Original Paper

Increasing the Effectiveness of a Physical Activity Smartphone Intervention With Positive Suggestions: Randomized Controlled Trial

Aleksandrina Skvortsova^{1,2}, PhD; Talia Cohen Rodrigues², MSc; David de Buissonjé², MSc; Tobias Kowatsch^{3,4}, PhD; Prabhakaran Santhanam³, MSc; Dieuwke S Veldhuijzen², PhD; Henriët van Middendorp², PhD; Andrea Evers^{2,5,6}, PhD

¹Department of Psychology, McGill University, Montreal, QC, Canada

²Health, Medical and Neuropsychology, Faculty of Social and Behavioural Sciences, Leiden, Netherlands

³Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

⁴Centre for Digital Health Interventions, Institute of Technology Management, University of St.Gallen, St.Gallen, Switzerland

⁵Department of Psychiatry, Leiden University Medical Center, Leiden, Netherlands

⁶Medical Delta, Leiden University, Technical University Delft and Erasmus University, Leiden, Delft, Rotterdam, Netherlands

Corresponding Author:

Aleksandrina Skvortsova, PhD

Department of Psychology

McGill University

1205 avenue du Docteur-Penfield

Montreal, QC, H3A 1B1

Canada

Phone: 1 4386303664

Email: a.skvortsova@fsw.leidenuniv.nl

Abstract

Background: eHealth interventions have the potential to increase the physical activity of users. However, their effectiveness varies, and they often have only short-term effects. A possible way of enhancing their effectiveness is to increase the positive outcome expectations of users by giving them positive suggestions regarding the effectiveness of the intervention. It has been shown that when individuals have positive expectations regarding various types of interventions, they tend to benefit from these interventions more.

Objective: The main objective of this web-based study is to investigate whether positive suggestions can change the expectations of participants regarding the effectiveness of a smartphone physical activity intervention and subsequently enhance the number of steps the participants take during the intervention. In addition, we study whether suggestions affect perceived app effectiveness, engagement with the app, self-reported vitality, and fatigue of the participants.

Methods: This study involved a 21-day fully automated physical activity intervention aimed at helping participants to walk more steps. The intervention was delivered via a smartphone-based app that delivered specific tasks to participants (eg, setting activity goals or looking for social support) and recorded their daily step count. Participants were randomized to either a positive suggestions group (69/133, 51.9%) or a control group (64/133, 48.1%). Positive suggestions emphasizing the effectiveness of the intervention were implemented in a web-based flyer sent to the participants before the intervention. Suggestions were repeated on days 8 and 15 of the intervention via the app.

Results: Participants significantly increased their daily step count from baseline compared with 21 days of the intervention ($t_{107}=-8.62$; $P<.001$) regardless of the suggestions. Participants in the positive suggestions group had more positive expectations regarding the app ($B=-1.61$, SE 0.47; $P<.001$) and higher expected engagement with the app ($B=3.80$, SE 0.63; $P<.001$) than the participants in the control group. No effects of suggestions on the step count ($B=-22.05$, SE 334.90; $P=.95$), perceived effectiveness of the app ($B=0.78$, SE 0.69; $P=.26$), engagement with the app ($B=0.78$, SE 0.75; $P=.29$), and vitality ($B=0.01$, SE 0.11; $P=.95$) were found. Positive suggestions decreased the fatigue of the participants during the 3 weeks of the intervention ($B=0.11$, SE 0.02; $P<.001$).

Conclusions: Although the suggestions did not affect the number of daily steps, they increased the positive expectations of the participants and decreased their fatigue. These results indicate that adding positive suggestions to eHealth physical activity interventions might be a promising way of influencing subjective but not objective outcomes of interventions. Future research should focus on finding ways of strengthening the suggestions, as they have the potential to boost the effectiveness of eHealth interventions.

Trial Registration: Open Science Framework 10.17605/OSF.IO/CWJES; <https://osf.io/cwjes>

(*J Med Internet Res* 2022;24(3):e32130) doi:[10.2196/32130](https://doi.org/10.2196/32130)

KEYWORDS

eHealth; mobile health; physical activity; walking; positive suggestions; outcome expectations; mobile phone

Introduction

Background

eHealth interventions use information and communications technology, such as mobile phones and computers, to improve or enable health care. The widespread availability of smartphones and computers makes it possible to provide eHealth interventions to broad populations without significant financial costs. A large fraction of eHealth interventions focus on behavioral change and implementing healthy lifestyle habits such as an increase in physical activity and diet change [1]. Accumulating literature shows that eHealth apps can increase physical activity in various groups of the population, including adolescents [2], working-age women [3], older people [4], patients with cardiovascular diseases [5], and survivors of cancer [6]. At the same time, multiple meta-analyses and literature reviews have demonstrated that the effectiveness of various eHealth interventions varies and that the effects of these interventions decrease over time [2,4,7,8]. Several characteristics of the interventions have been proposed to increase their effectiveness: user-friendly design, real-time feedback, and health professional involvement [1].

A possible way of increasing the effectiveness of eHealth interventions, which has not yet been investigated, is to manipulate the expectations of users regarding the intervention. This can be accomplished by providing users with positive suggestions that emphasize the effectiveness of the intervention. Positive outcome expectations are one of the primary mechanisms of placebo effects. A lot of research has been conducted on the effects of positive suggestions on the outcomes of various interventions [9-11]. For example, it has been demonstrated that optimizing the expectations of patients undergoing bypass surgery leads to lower disability in these patients 6 months after surgery [10]. Placebo effects induced by enhancing positive expectations of patients have also been found in pain, itch, depression, fatigue, and nausea [12-15]. In addition to changing the expectations of users, positive suggestions about the intervention can also increase the perceived credibility of the intervention and adherence of the users to the intervention [9,16].

Although placebo effects induced by positive suggestions have been extensively investigated in various areas of health care, studies that look at the potential of using placebo effects in eHealth remain scarce. A recent study has used several types of positive suggestions to change the expectations of participants

regarding a smartphone-delivered placebo intervention aimed at improving mood [17] and has found that expectations about the app and its credibility decreased during the 20 days of the placebo intervention. However, this decrease was less prominent in a combined suggestions condition in which participants were informed about the positive effects of the app before the start of the intervention and provided with positive feedback during the intervention [17]. Although this previous study aimed to increase positive expectations regarding a placebo intervention, no study to date has aimed to increase the effectiveness of an active eHealth intervention by changing the expectations of participants.

Objectives

This study investigates whether positive suggestions can influence the expectations of participants and increase the effectiveness of a physical activity smartphone intervention. We look at the effects of positive suggestions given before and during the intervention on several outcomes. The primary outcomes are (1) expectations of the participants regarding the effectiveness of the app and (2) the number of steps participants took during the intervention. The secondary outcomes are the perceived effectiveness of the app, engagement with the app, and vitality and fatigue of participants during the intervention.

Methods

Ethics Approval

This study was approved by the psychology research ethics committee of Leiden University (2020-09-14-AWMEvers-V2-2625). The study protocol was preregistered on Open Science Framework [18].

Study Design

A randomized, between-subjects study design was used. Participants were randomly allocated to one of the two conditions: (1) positive suggestions group (intervention with positive suggestions) and (2) control group (intervention without suggestions). A random number generator was used to block randomize participants to their condition with a block size of 6.

The study was web based, and all measurements were performed via the internet. During the intervention, participants did not have direct contact with the researchers; however, in case they had questions, they could get in touch with the research team via email.

Participants

Healthy participants aged between 18 and 40 years were recruited for this study. Recruitment was conducted across the campus of Leiden University and using social media such as Facebook and WhatsApp groups of students. Inclusion criteria were the ability to sufficiently understand, read, and write English; ability to use smartphones and the internet; being in possession of a smartphone; and being willing to increase physical activity. Participants with medical conditions that could hinder a normal physical activity pattern (eg, joint problems or heart disease) were excluded from this study. Participants received €10 (US \$11.43) or 8 study credit points for participating in the study in case they completed the whole intervention, €6 (US \$6.86) or 6 study credits in case they completed 2 weeks, and €3 (US \$3.43) or 3 credits if they completed the first week of the intervention. The reward was given after the end of the study.

As no research so far has been performed on the effects of positive suggestions on the step count during an eHealth intervention, a study on the effects of positive suggestions on physical performance (weightlifting exercises) was chosen for the sample size calculation [19]. The power calculation, conducted with G*Power 3.1 [20], indicated that to detect a difference between positive suggestions and a control group using analysis of covariance, with an estimated effect size of Cohen $f=0.47$ [19], a critical α level of $\alpha=.05$, and a power of $\beta=.95$, 31 participants per group, that is, 62 participants in total, would be needed. In a similar study by our laboratory, which used the same eHealth intervention in healthy volunteers, the dropout rate from the moment of recruitment until the end of the study was 33.6% (46/140). Therefore, we aimed to recruit 93 participants in this study.

Procedure

The study was advertised as a study testing a mobile phone physical activity intervention. Participants interested in the study were sent an information letter with the details about the study and were asked to digitally sign an informed consent form on the Qualtrics platform (Qualtrics International Inc). Participants were asked to ensure that they had either Google Fit or Apple Health apps installed on their phones at least 1 week before the

start of the intervention. The step count data from these apps from the week preceding the intervention were retrieved by the study app and used as a baseline measurement of the average number of steps taken by the participants.

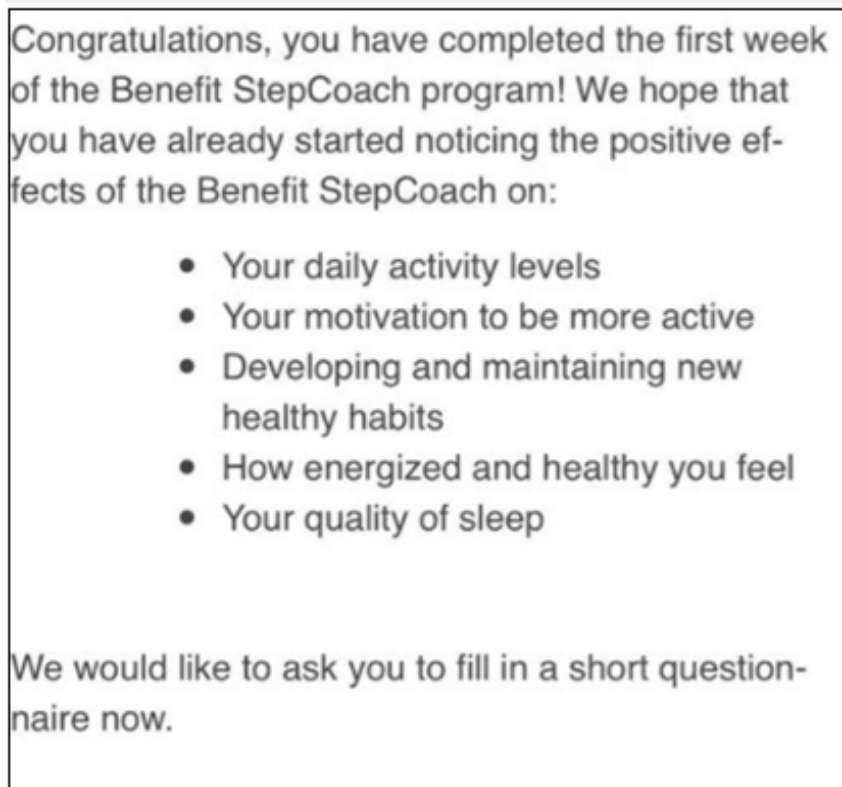
Participants were sent a download link on Apple's App Store and Google Play Store 1 week before the start of the intervention. After downloading the app, they were asked to fill in several baseline web-based questionnaires measuring their expectations about the app, vitality and fatigue they experienced during the past 2 weeks, motivation to exercise, anxiety, and expected engagement with the app. At this point, they were also asked to give permission to retrieve their step count from Google Fit or Apple Health and send them push notifications.

On day 1 of the intervention, participants were randomly allocated to one of the following intervention conditions: intervention with positive suggestions or control intervention (without suggestions). On the morning of day 1 of the intervention, participants in the positive suggestions group received a web-based flyer with positive suggestions (Figure 1, left). Participants in the control group received a similar flyer describing the technical details of the study app (Figure 1, right). Participants were asked to read the flyers carefully, as the information given on them would be needed during the intervention. After reading the flyers, participants were asked to fill in a questionnaire measuring their expectations regarding the intervention. Every day of the intervention, at 9 AM, participants received a push notification from the study app. In addition, the study app retrieved the step data from Google Fit or Apple Health and tracked the number of steps participants took daily. On days 8 and 15 of the intervention, participants in the positive suggestions group were given brief booster suggestions regarding the intervention (Figure 2), and participants in both groups were asked to fill in short questionnaires about how effective and engaging they found the intervention. After the last day of the intervention, participants were asked to fill in several questionnaires that measured their perceived effectiveness of the intervention, engagement with the app, vitality, and fatigue. The flowchart with the steps of the study is presented in Multimedia Appendix 1.

Figure 1. The flyers with positive (left) and control (right) suggestions.



Figure 2. The screenshot of the short suggestions repeated on days 8 and 15 of the intervention.



Experimental Interventions

Study Conditions

Positive Suggestions Group

In the positive suggestions group, before the start of the intervention, participants received a leaflet with information regarding the effectiveness of the intervention. It emphasized that the study app has been shown to successfully motivate people to be more physically active and listed various positive effects of using the app. The flyer with positive suggestions is presented in [Figure 1](#) (left).

In addition, the suggestions were implemented in the study app and given to the participants again (booster suggestions) on days 8 and 15 of the intervention (end of the first and second weeks of the intervention). Only 2 very short boosters were given to avoid introducing too much difference in the amount of information that the control and experimental groups had to read. A screenshot of the booster suggestions is presented in [Figure 2](#).

Control Group

In the control group, participants received a leaflet with the technical details about the study app. The information was similar in length to the information in the positive suggestions group but did not aim to influence the expectations of participants regarding the effectiveness of the intervention. The control leaflet is presented in [Figure 1](#) (right).

The Study App and Intervention

The study app was developed with MobileCoach [21,22], an open-source software platform for smartphone-based and

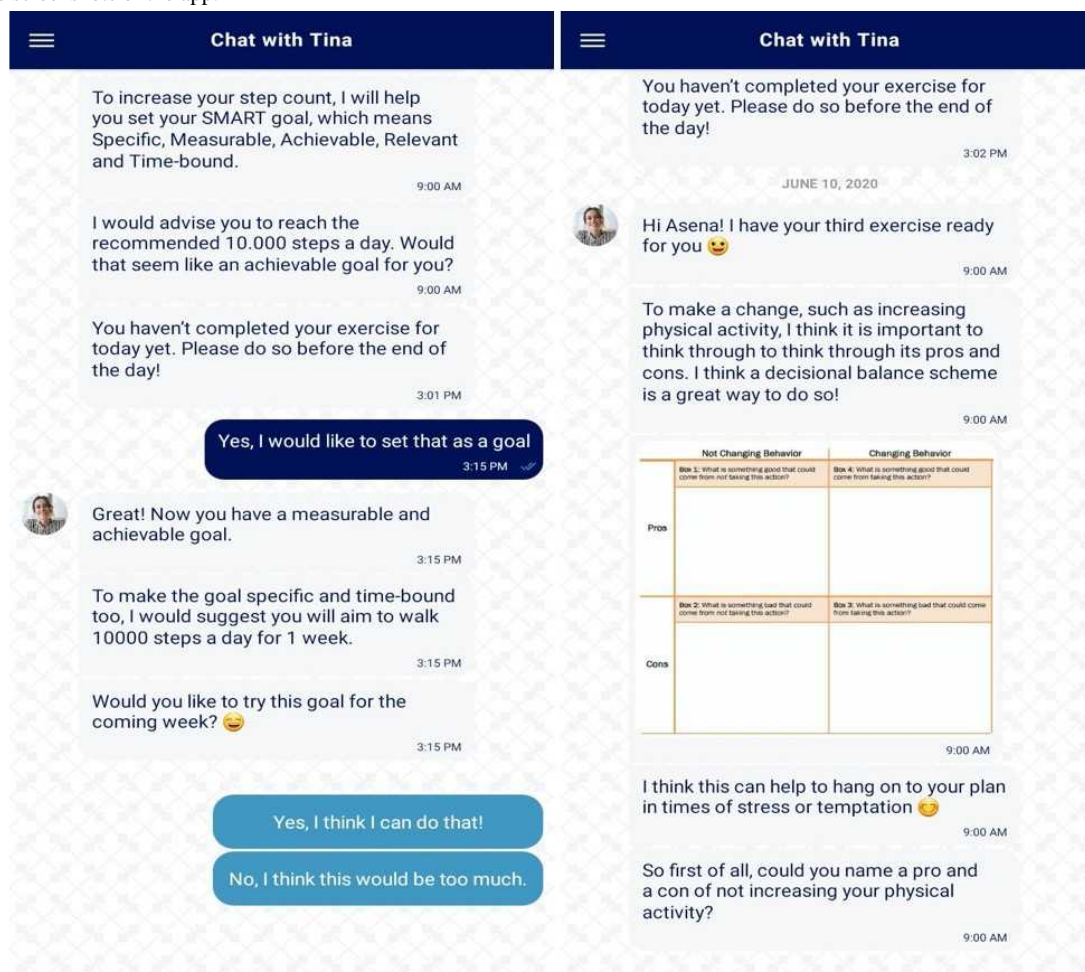
chatbot-delivered behavioral health interventions and ecological momentary assessments. MobileCoach-based interventions [23,24] have been used in various studies for, for example, stress management [25], personality change [26], promotion of health literacy [23], or physical activity [24]. The graphical user interface of the app is similar to WhatsApp or other messaging apps to leverage participants' expertise in already-existing and well-known communication and interaction paradigms. The app connected to the Apple Health or Google Fit apps on the smartphone of a user and retrieved the step count from these apps every day of the intervention. Moreover, the app had integrated LimeSurvey questionnaires (LimeSurvey GmbH) that allowed the questionnaires to be sent to the users directly through the study app. The MobileCoach-based app of this study aimed to increase the physical activity of its users by increasing their daily step count. The app was tested in other research projects of our laboratory. No bugs or inconsistencies were found in the intervention during the trial. The step data of the participants who finished the intervention were fully available. The step data of participants who stopped with the intervention and did not open the app anymore were not available from the moment they stopped opening the app.

The study intervention included onboarding and 21 days of active intervention. During the 21 intervention days, every morning, participants received several messages from the chatbot of the app, framed as a mobile step coach. These messages contained short psychological exercises that helped users set reasonable activity goals and think about reasons for participating in the intervention, possible barriers, and ways to overcome these barriers. Similar to other MobileCoach-based interventions [23,24], participants could communicate with the chatbot and type their responses in the app or use predefined

answer options. The app used both predefined scripts and predefined answer options for maximum safety and traceability of the intervention progress. The conversational turns with the chatbot included human cues, for example, a picture of the coach, humor, and references to the personal life of the coach, to make the chatbot more human-like [27,28] and trigger a working alliance [23,29,30], which is robustly linked to

intervention outcomes [31-33]. A screenshot of the app is presented in Figure 3. As the chatbot was not preprogrammed to converse on topics not related to the intervention, it was not able to reply to the spontaneous questions of participants. If participants needed assistance with the app or had any questions, they could get in touch with the researchers via email.

Figure 3. The screenshots of the app.



The intervention was based on the Transtheoretical Model of Health Behavior Change [34]. This model views behavioral change as an upward spiraling process involving progress through five stages: precontemplation, contemplation, preparation, action, and maintenance. Each exercise of the intervention targeted one of the stages. In addition, several behavior change techniques were incorporated into the intervention, such as prompts, cues, information about health consequences, review of goals, and social rewards [35]. An overview of the intervention is presented in Multimedia Appendix 2 (including the stages of the health behavior change targeted by the exercises).

Measurements and Questionnaires

Demographic Characteristics

Participants were asked about their age, gender, nationality, and height and weight. BMI was calculated based on the height and weight of participants using the following formula: weight (kg) divided by height (m²).

Primary Outcomes

Steps were measured by Google Fit or Apple Health on the phone of the participants and retrieved from these apps into the study app. Steps were measured for 1 week before the start of the intervention and for the 21 days of the intervention. The mean number of steps from the week before the start of the intervention was used as a baseline step measure. The number of steps on each day of the intervention was used in the analyses as the main outcome measure.

Expectations regarding the effectiveness of the app were measured with six statements created for this study: (1) the app will help me become more physically active; (2) the app will motivate me to increase my step count; (3) the app will help me enjoy moving more; (4) the app will help me develop and maintain healthy habits; (5) using the app will make me feel more energized; and (6) using the app will make me sleep better. Participants were asked to evaluate these statements using a 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree). Expectations were measured

twice—at baseline before the flyer was given and after participants read the flyers with positive or control suggestions. The total score of the 6 statements was used in the analysis. The score could range between 6 and 30, with a higher score indicating more positive expectations. The internal consistencies of both baseline ($\alpha=.79$) and day 1 ($\alpha=.80$) measures were acceptable and good, respectively.

Secondary Outcomes

Perceived app effectiveness was measured with six statements created for this study that mirrored the questions about expectations: (1) the app helps me become more physically active; (2) the app motivates me to increase my step count; (3) the app helps me enjoy moving more; (4) the app helps me develop and maintain healthy habits; (5) using the app makes me feel more energized; and (6) using the app makes me sleep better. Participants were asked to evaluate these statements using a 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree). The total score of the 6 statements was used in the analysis. The score could range between 6 and 30, with higher scores indicating more positive expectations. Perceived effectiveness was measured three times: on days 8 and 15 and after the last day of the intervention. The internal consistencies of day 8 ($\alpha=.86$) and day 15 ($\alpha=.87$) measures were good and excellent ($\alpha=.90$), respectively.

(Expected) Engagement with the app was measured with the TWente Engagement with EHealth Technologies Scale [36]. Two versions of the questionnaire were used: 1 measuring expected engagement and 1 measuring current engagement. The questionnaire contained 9 statements. The expected engagement questionnaire, measured at baseline and on day 1, asked users about how engaging they thought they would find the app. The engagement questionnaire, measured on days 8 and 15 and at the end of the intervention, asked about how engaging the users found the intervention. Participants were asked to rate the statements using a 5-point Likert scale (1=strongly disagree, 2=disagree, 3=neutral, 4=agree, and 5=strongly agree). The total score could range between 9 and 45, with higher scores on the questionnaire indicating higher engagement with the app. The internal consistency of the baseline measure was questionable ($\alpha=.69$); however, the day 1 measure was of acceptable consistency ($\alpha=.79$), and the day 8, day 15, and last day measures had good internal consistency ($\alpha=.83$, $.85$, and $.83$, respectively).

Vitality was measured using the Subjective Vitality Scale [37]. The Subjective Vitality Scale comprises 7 statements and measures the subjective feeling of being alive and alert during the past 2 weeks. Participants were asked to rate how true they found the statements on a 7-point Likert scale (1=not at all; 7=very true). The score could range between 7 and 49; higher scores on the composite scale could be interpreted as representing higher self-reported vitality. Vitality was measured at baseline before the start of the intervention and after the last day of the intervention.

Fatigue was measured using the Checklist Individual Strength [38]. The Checklist Individual Strength comprises 20 statements that measure four dimensions of fatigue experienced during the past two weeks: fatigue severity, concentration problems,

reduced motivation, and activity. Participants were asked to rate how true they found the statements on a 7-point Likert scale (1=yes, that is true; 7=no, that is not true). The score ranged between 20 and 140, and higher scores could be interpreted as higher fatigue. Fatigue was measured at baseline before the start of the intervention and after the last day of the intervention.

The motivation to exercise was measured using the Behavioral Regulation in Exercise Questionnaire-2 [39]. The scale comprises 19 items. Participants were asked to indicate how true they found the statements on a 7-point Likert scale (1=not true for me; 7=very true for me). The questionnaire comprises five subscales: amotivation, external regulation, introjected regulation, identified regulation, and intrinsic regulation. Higher scores indicate higher motivation. The motivation to exercise was measured once at the baseline.

Anxiety was measured using the Generalized Anxiety Disorder 7-item scale [40]. Participants were asked to indicate how often they had been bothered by a list of problems in the past 2 weeks on a 4-point Likert scale (0=not at all; 3=nearly every day). The score ranged between 0 and 21, with higher scores indicating higher levels of anxiety. Anxiety was measured once at baseline.

Statistical Analysis

Statistical analyses were performed using RStudio (version 1.1.447; R version 4.0.4). The data and the R code are made available on the web [18]. All tests were performed 2-tailed.

To compare groups on baseline characteristics (age, BMI, steps during the week before the start of the intervention, baseline expectations, fatigue, vitality, baseline expected engagement with the MobileCoach, and motivation to exercise), independent-sample 2-tailed *t* tests or nonparametric Wilcoxon tests (in case of violations of assumptions) were used. To examine whether participants enrolled in the first and second rounds of the recruitment differed on the baseline measurements and the study outcomes (baseline steps, baseline and postintervention vitality and fatigue, steps during the intervention, and postintervention vitality and fatigue), independent-sample *t* tests or nonparametric Wilcoxon tests (in case of violations of assumptions) were used.

To investigate whether the MobileCoach intervention had an effect on the number of steps participants took daily, we compared the mean of daily steps participants took in the week before the intervention with the mean of daily steps during the intervention with a paired-sample *t* test.

The lmer function of the nlme package in R (R Core Team, 2013) was used for the linear mixed effects model analyses to test the main hypotheses of the study. The multilevel structure of the data was defined by *day* (level 1) nested in *participants* (level 2). Parameters were estimated using the full maximum-likelihood procedure. In all models, the intercept was allowed to vary randomly across participants. Random slopes did not improve the fit of the models; therefore, they were removed from the final analysis. The effect sizes (Cohen *d*) of all linear mixed effects models were calculated using the EMAtools package. Cohen *d*=0.2 was interpreted as a small effect size, Cohen *d*=0.5 as a medium effect size, and Cohen *d*=0.8 as a large effect size [41].

To investigate whether the suggestions had an effect on the expectations of the participants regarding the effectiveness of the intervention, we used a linear mixed effects model. The expectations of day 1 (postsuggestions) were the dependent variable, and the independent variables were group and baseline expectations (presuggestions).

To examine whether the suggestions had an effect on the steps taken during the intervention, we used a linear mixed effects model approach. Steps per each of the 21 days of the intervention were used as the dependent variable, and group, baseline steps (to control for the individual baseline differences in walking), and day were the independent variables.

To examine whether the suggestions had an effect on the perceived effectiveness of the intervention, a linear mixed effects model was used with the effectiveness of the intervention as the dependent variable and group and day as independent variables.

To examine whether the suggestions had an effect on the expected engagement with the app, a linear mixed effects model was used with postsuggestions expected engagement as the dependent variable and group and presuggestions expected engagement as independent variables. To investigate whether suggestions had an effect on engagement with the app during the intervention, a linear mixed effects model was used with engagement as the dependent variable and day and group as independent variables.

To examine whether the suggestions had an effect on vitality and fatigue, 2 separate linear mixed effects models were used with vitality or fatigue during the intervention as the dependent variable and group and baseline vitality or fatigue as independent variables.

To examine whether postsuggestion expectations had an effect on the dependent variables, several similar models were created with postsuggestion expectations instead of the group as an independent variable and step count, perceived app effectiveness, (expected) engagement, vitality, and fatigue as dependent variables.

Results

Participants

In the first round of the study, 93 participants were recruited between November 2020 and December 2020. As the inclusion criteria were presented in the advertisement, no participants had to be excluded based on not meeting the inclusion criteria. However, the dropout rate was larger than expected, and full data of only 39.8% (37/93) participants were available after the first round of data collection. Therefore, the intervention was performed again in February 2021, with 45 more participants recruited into the study to reach the intended sample size. No follow-up was conducted with the participants who stopped prematurely because of the web-based nature of the study. Therefore, the reasons why the participants dropped out remain unknown.

In total, 138 participants were randomized between 2 groups; 133 (96.4%) participants started the study, and the full data of 79 (57.2%) participants were available. The flowchart of the participants included in each step and dropouts is presented in [Multimedia Appendix 3](#). An overview of the missing data is presented in [Multimedia Appendix 4](#). All available data were included in the analyses, and the number of participants included in each analysis is presented in the results.

Demographics and Baseline Characteristics

Of the 133 participants, 69 (51.9%) participants (n=55, 80% women; n=13, 19% men; and n=1, 1% other) were assigned to the positive suggestions group and 64 (48.1%; n=56, 88% women and n=8, 13% men) to the control group. The mean age of the sample was 23.3 (SD 6.1) years. An overview of the baseline characteristics across the groups with comparison tests is presented in [Table 1](#). No differences in characteristics were found between any of the groups. No differences were found in the baseline measurements and the outcomes of the study between the participants who took part in the first and second rounds of the recruitment.

Table 1. Mean scores of the baseline variables per group and the comparison statistics (N=133).

Variable	Positive suggestions group (n=69)		Control group (n=64)		<i>t</i> test (<i>df</i>)	Wilcoxon test ^a	<i>P</i> value
	Value, mean (SD)	Value, n (%)	Value, mean (SD)	Value, n (%)			
Age (years)	24.00 (6.79)	69 (100)	22.45 (5.26)	64 (100)	— ^a	1939	.22
BMI (kg/m ²)	23.08 (4.52)	66 (96)	22.71 (3.38)	61 (95)	—	2004	.97
Steps	4060.47 (2271.2)	55 (80)	4172.77 (2576.11)	53 (83)	—	1457	.99
Expectations about app effectiveness	21.29 (2.87)	69 (100)	21.05 (3.12)	64 (100)	—	2171	.87
Expected engagement with the study app	30.06 (2.75)	69 (100)	30.20 (2.86)	64 (100)	0.30 (131)	—	.77
Vitality	4.25 (0.88)	69 (100)	4.28 (0.87)	64 (100)	0.25 (131)	—	.81
Fatigue	30.12 (9.61)	69 (100)	30.98 (8.95)	64 (100)	0.54 (131)	—	.59
Anxiety	5.49 (3.97)	69 (100)	6.05 (4.7)	64 (100)	—	2279	.75
Motivation to exercise							
Amotivation	5.61 (2.49)	69 (100)	5.62 (2.57)	64 (100)	—	2170	.85
External	6.04 (2.65)	69 (100)	6.23 (3.16)	64 (100)	—	2178	.89
Introjected	8.54 (3.37)	69 (100)	9.45 (3.27)	64 (100)	—	2508	.18
Identified	14.46 (3.72)	69 (100)	14.59 (3.82)	64 (100)	—	2255	.83
Intrinsic	13.94 (4.6)	69 (100)	13.59 (4.36)	64 (100)	—	2256	.83

^aWilcoxon test coefficient was used in case a variable was not normally distributed.

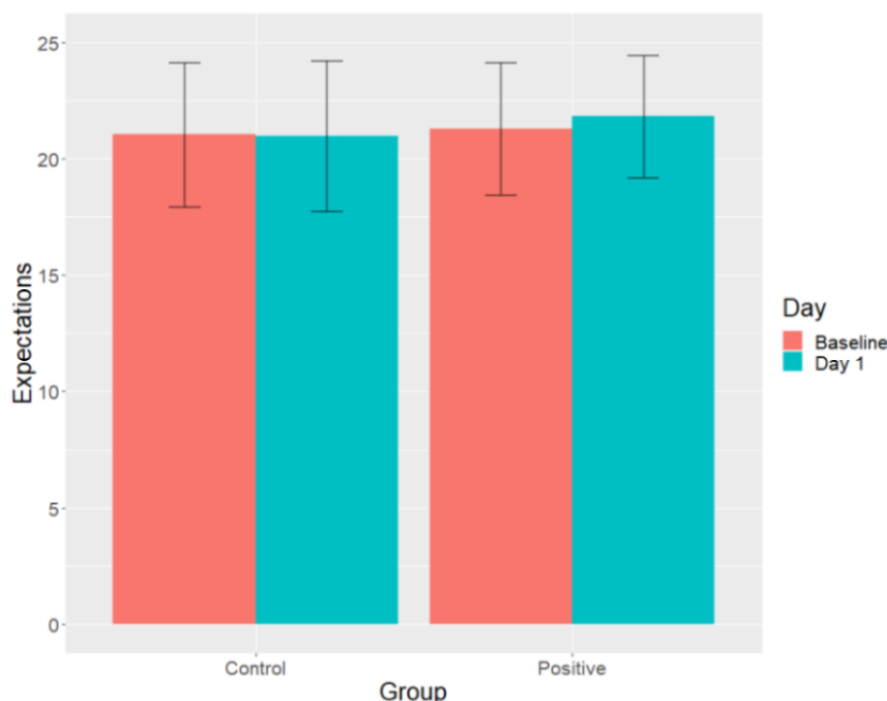
Primary Analyses

Expectations

Expectations significantly differed between groups, with a medium effect size (B=-1.61, SE 0.47; *t*₁₂₄=-3.40; *P*<.001;

Cohen *d*=0.61; 125/133, 94%) when controlling for the presuggestion expectations: participants in the positive suggestions group (mean 21.82, SD 2.64) expected the app to be more effective than participants in the control group (mean 20.98, SD 3.23). The mean expectations for each group are presented in Figure 4.

Figure 4. Mean expectations scores per group at baseline and after the flyer with suggestions was sent to participants on day 1. The error bars indicate SDs.



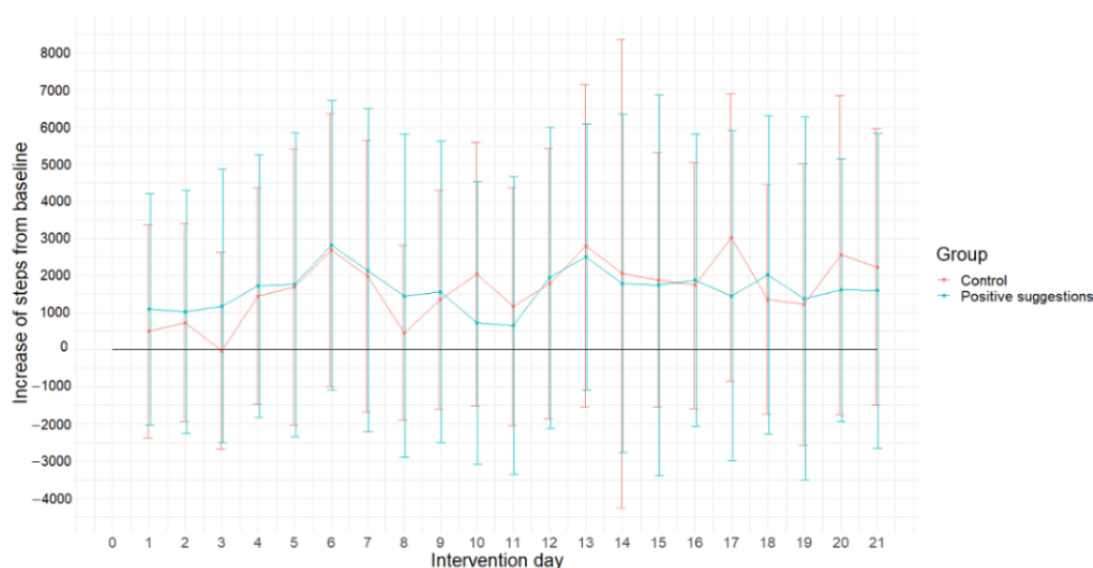
Steps

Irrespective of the group, participants significantly increased the number of daily steps during the intervention (mean 5689.6, SD 2718.3) compared with the mean number of daily steps during the week before the intervention (mean 4115.5, SD 2414.9; $t_{107}=-8.62$; $P<.001$; 108/133, 81.2%).

The difference between the steps performed on each day of the intervention and the mean number of steps from the baseline week is shown in Figure 5. The multilevel model with the number of steps during the intervention as a dependent variable; baseline steps, day, and group as predictors; and a random

intercept (108/133, 81.2%) demonstrated no significant effect of the group on the number of steps ($B=-22.05$, SE 334.90; $t_{105}=-0.07$; $P=.95$; Cohen $d=0.01$). The model showed that the variable day significantly predicted the number of steps: participants in both groups increased the number of steps they took daily during the intervention ($B=37.47$, SE 12.51; $t_{2009}=2.99$; $P=.003$). However, the effect size of the day as a predictor was negligible (Cohen $d=0.13$). Baseline steps also significantly predicted the steps taken during the intervention with a large effect size ($B=0.71$, SE 0.07; $t_{105}=10.48$; $P<.001$; Cohen $d=1.97$): participants who took more steps in the baseline week also took more steps during the intervention.

Figure 5. The difference between the mean number of steps taken during the intervention per day and the baseline mean steps per group. The error bars indicate SDs.



The model with postsuggestion expectations instead of the group (108/133, 81.2%) demonstrated that expectations did not influence the number of steps taken during the intervention ($B=-44.06$, SE 59.68; $t_{105}=-0.74$; $P=.46$; Cohen $d=0.14$). The effects of the day ($B=37.56$, SE 12.51; $t_{2009}=3$; $P=.003$; Cohen $d=0.13$) and baseline steps ($B=0.70$, SE 0.07; $t_{105}=10.08$; $P<.001$; Cohen $d=1.97$) were also significant in this model.

Secondary Analyses

Perceived App Effectiveness

The reported effectiveness of the app is presented in Multimedia Appendix 5.

The multilevel model with the app effectiveness as a dependent variable, group and day number as predictors, and a random intercept (122/133, 91.7%) demonstrated that the group did not influence the perceived app effectiveness ($B=0.78$, SE 0.69; $t_{120}=1.12$; $P=.26$; Cohen $d=0.21$). The day variable had an effect on the effectiveness: participants reported an increase in the effectiveness of the app during the intervention ($B=0.06$, SE 0.02; $t_{210}=2.81$; $P=.005$; Cohen $d=0.39$).

(Expected) Engagement With the App

The multilevel model with group and baseline (presuggestions) expected engagement as predictors and random intercept varying

across participants (127/133, 95.5%) demonstrated that the group had a significant effect on the postsuggestions expected engagement ($B=3.80$, SE 0.63; $t_{124}=6.04$; $P<.001$; Cohen $d=1.09$) when controlling for the presuggestions expected engagement: participants in the positive suggestions group (mean 31.62, SD 3.31) expected the app to be more engaging than participants in the control group (mean 30.41, SD 3.26).

Another model with engagement as a dependent variable, and group (122/133, 91.7%) and day number as predictors demonstrated that neither group ($B=0.78$, SE 0.75; $t_{120}=1.06$; $P=.29$; Cohen $d=0.19$) nor day ($B=0.042$, SE 0.03; $t_{210}=1.60$; $P=.11$; Cohen $d=0.22$) had an effect on the reported engagement.

Vitality

The multilevel model with vitality during the intervention as a dependent variable, group and baseline vitality as predictors, and a random intercept (117/133, 88%) demonstrated that the group did not have an effect on vitality during the intervention ($B=0.01$, SE 0.11; $t_{114}=-0.07$; $P=.95$; Cohen $d=0.01$). Baseline vitality was positively associated with vitality during the intervention ($B=0.20$, SE 0.02; $t_{114}=11.11$; $P<.001$; Cohen $d=2.08$).

Fatigue

The multilevel model with fatigue during the intervention as a dependent variable and group and baseline fatigue as predictors (133/133, 100%) demonstrated that the group significantly predicted fatigue ($B=3.09$, $SE\ 0.59$; $t_{130}=5.22$; $P<.001$; Cohen $d=0.92$): participants in the positive suggestions group (mean 23.49, $SD\ 12.47$) reported lower fatigue than participants in the control group (mean 24.94, $SD\ 11.62$). Baseline fatigue positively predicted fatigue during the intervention ($B=0.11$, $SE\ 0.02$; $t_{130}=6.68$; $P<.001$; Cohen $d=1.17$).

Postsuggestion Expectations

Postsuggestions expectations significantly influenced the perceived app effectiveness ($B=0.62$, $SE\ 0.10$; $t_{119}=6.13$; $P<.001$; Cohen $d=1.12$; 121/133, 91%): the higher expectations participants had about the app on day 1, the higher the app effectiveness they reported during the intervention. Postsuggestion expectations also significantly, and with a large effect size, influenced the vitality of participants ($B=0.02$, $SE\ 0.004$; $t_{114}=5.09$; $P<.001$; Cohen $d=0.95$; 117/133, 88%): higher expectations were linked to higher vitality during the intervention. The postsuggestion expectations had no effect on fatigue ($B=-0.04$, $SE\ 0.02$; $t_{124}=-1.79$; $P=.08$; Cohen $d=0.32$, 127/133, 95.5%).

Discussion

Principal Findings

The aim of this study was to investigate whether it is possible to improve the effectiveness of a physical activity smartphone intervention through positive suggestions. We demonstrated that positive suggestions affected the expectations of participants regarding the effectiveness of the app; however, they were not effective enough to affect the main outcome; that is, daily step count. In addition, participants in the positive suggestions group reported higher expected engagement with the app and lower fatigue during the intervention than the participants in the control group.

The smartphone intervention was effective in helping people increase their physical activity. Participants in both groups walked on average 1500 ($SD\ 2517$) steps more during the intervention than before the intervention. Moreover, there was some indication that participants increased their daily steps as the intervention progressed; however, despite the significant effect, the effect size was negligible. This result was expected as the intervention included the phases of the Transtheoretical Model of Health Behavior Change that starts with precontemplation and contemplation about the behavior change and ends with the maintenance of new health behavior [34]. A number of studies demonstrated that physical activity interventions based on the Transtheoretical Model of Health Behavior Change could efficiently increase physical activity in various populations [42-44]. In addition, multiple other smartphone interventions based on different theoretical approaches were shown to increase the physical activity of app users [2,3,8]. Therefore, our results are in line with the literature showing that a psychological smartphone intervention can be

successful in helping people increase their physical activity, at least in the short term.

Two outcomes were considered primary in this study: expectations of participants regarding the effectiveness of the intervention and the number of steps taken during the intervention. We hypothesized that positive suggestions would change the expectations of participants, and these increased positive expectations, in turn, would lead to a better outcome of the intervention—an increase in the daily step count. This model is often described in the literature as a working mechanism of placebo effects: positive suggestions induce a change in expectations, and positive expectations lead to better treatment outcomes [16]. The first part of our primary hypothesis was supported by the data: participants in the positive suggestions group indeed reported expecting better effectiveness of the app than the control group. Although the change in the expectations score was quite small, the effect size was shown to be medium. This result is in line with the broad literature on placebo effects: multiple studies demonstrated that positive suggestions given by a health care professional, or an experimenter, can change the expectations of people regarding various types of interventions—in pain [45,46], itch [47,48], nausea [49,50], and many other symptoms [13,15,51].

Despite the effectiveness of the expectancy manipulation, changing the expectations of participants was not enough to influence their daily step count: no difference was found between the groups in their number of steps during the 21 days of the intervention. Moreover, expectations regarding app effectiveness were not predictive of the number of steps participants made, which may explain why an increase in expectations did not affect the step count. These results do not support our main hypothesis and contradict the literature on treatment expectations [52]. The reason for this contradiction can be that our study differs from other studies that used positive suggestions. Most of the research applied positive suggestions focused on symptom-reducing interventions, such as interventions decreasing pain [45,46], itch [53], or nausea [54] or improving mood [55,56]. The intervention that we used in our study aimed to change the behavioral habits of people and required active action from users. Different mechanisms might be involved in interventions that relieve subjective symptoms, such as pain, and interventions that target to change behavior. We can speculate that expecting an overall positive change is not specific enough to improve the outcomes of such interventions and affect their behavior. According to the self-efficacy theory of Bandura [57,58], self-efficacy beliefs have to be very specific to successfully influence behavior, and possibly this specificity was lacking in the suggestions we gave. Generic positive suggestions about an intervention might be more powerful for some subjective outcomes, such as fatigue, which was influenced by positive suggestions in our study. Similar results were found in previous research aimed at influencing food choices by positive suggestions: suggestions influenced subjective food preferences but not actual behavior [59,60]. Our results confirm the results of multiple studies that have demonstrated that fatigue can be decreased by positive suggestions regarding various placebo treatments [19,61].

Another explanation for the failure to confirm the main hypothesis for step count could be the fact that the suggestions given in this study were quite minor. Participants received a flyer with suggestions at the beginning of the study, and the suggestions were briefly repeated on days 8 and 15 of the intervention. It is possible that if the suggestions were more extensive and implemented in each element of the intervention itself, they might have affected the physical activity of the participants.

Finally, the possibility exists that the suggestions did not influence the number of steps but some other parameters that, in turn, could have affected fatigue. The app was not able to measure the pace of walking or the frequency of walking of participants. If positive suggestions affected the walking pace and frequency, these increases in physical activity could have caused the experimental group to experience less fatigue during the intervention.

Limitations

Several limitations of this study have to be addressed. First, the participants of this study were young and healthy university students. This group is very familiar and comfortable with using mobile phone apps. Therefore, the effects found in this study might not be generalizable to older and less technically experienced populations. It can be that older and less technically experienced people would find it harder to use the app, and therefore, the effects of the app could be smaller in this group of people. At the same time, our main goal was to investigate the effects of positive suggestions on app effectiveness, and there is no particular reason for assuming that computer literacy would influence these effects. Second, several other control groups could have been included in the study to enable us to interpret the results more fully. For example, this study did not include a control group that did not participate in the intervention. This fact makes it impossible to pinpoint whether the intervention was the reason for the increase in physical activity of participants. However, as the main aim was to investigate the effect of positive suggestions, we omitted this experimental condition. In addition, a comparison was made between a group that received a flyer with positive suggestions

and a group that received a flyer with technical details. The possibility remains that the fact of presenting users technical versus nontechnical information influenced the study results on top of the presentation of positive versus neutral information. Future research should investigate whether nontechnical neutral information about eHealth apps differently influences their effectiveness. Another limitation of the study is that no follow-up was performed. It remains unknown whether participants would maintain the new habits that they developed during the intervention or the positive effects would disappear quickly after the intervention's end. Moreover, a selection bias may have been present in this study. It was advertised as a study aimed at increasing physical activity, which might have attracted participants who were already interested in changing their lifestyles. It would be interesting to see how effective the suggestions would be in less motivated or less healthy participants, particularly as a recent meta-analysis of the effects of physical activity eHealth interventions demonstrated that clinical and at-risk groups benefit from such interventions more than healthy volunteers [8]. Another possible limitation of this study is that social desirability might have played a role in the effects we have found. Studies show that similar social desirability patterns occur in interactions with conversational agents, especially when the conversational agent is more human-like [62,63]. The study app used human cues to establish a better working alliance with the participants. However, it could have triggered the participants to give more socially desirable answers on subjective self-report measures.

Conclusions

In summary, we have demonstrated that short positive suggestions regarding the effectiveness of a mobile health app can affect the expectations of users and decrease their fatigue. Although no effect was found on the main outcome—the step count—these results indicate that eHealth interventions might benefit from adding positive suggestions to the description of the apps. Suggestions are an easy and effective way of giving an extra boost to the effectiveness of eHealth interventions, which does not require extra time or investment. Future research should focus on optimizing such suggestions to achieve actual behavior changes.

Acknowledgments

The authors would like to acknowledge bachelor's students Ivana Dumucic, Vanessa Eastick, Emma Franke, Nicola Hughes, Katja Salauyova, Ines Sánchez-Saura, and Rasa Urbanaviciutes for their help with data collection. The authors would also like to thank the members of BENEFIT Consortium. This work was supported by an NWO Vici Grant (016.Vici.170.152) granted to AE, the Dutch Heart Foundation, and ZonMw (The Netherlands Organization for Health Research and Development). The funders had no role in data acquisition, data analyses, or the content of the manuscript.

Authors' Contributions

AS, DSV, HvM, and AE developed the idea of the study. TCR, TK, and PS designed and implemented the MobileCoach-based intervention. AS collected the data and performed the statistical analysis. AS wrote the first version of the manuscript with input from all authors. AE supervised the entire project.

Conflicts of Interest

PS and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at ETH Zürich and the Institute of Technology Management at the University of St Gallen, which

is funded in part by the Swiss health insurer CSS. TK is also a cofounder of Pathmate Technologies, a university spinoff company that creates and delivers digital clinical pathways and has used the open-source MobileCoach platform for that purpose as well. However, neither CSS nor Pathmate Technologies were involved in the design, data analysis, or results reported in this study.

Multimedia Appendix 1

The flowchart of the study methods.

[[PNG File , 39 KB - jmir_v24i3e32130_app1.png](#)]

Multimedia Appendix 2

Overview of the intervention.

[[PNG File , 1336 KB - jmir_v24i3e32130_app2.png](#)]

Multimedia Appendix 3

Study flowchart depicting the number of participants in each step of the study.

[[PNG File , 36 KB - jmir_v24i3e32130_app3.png](#)]

Multimedia Appendix 4

Overview of the missing data per outcome variable.

[[PNG File , 35 KB - jmir_v24i3e32130_app4.png](#)]

Multimedia Appendix 5

Mean perceived effectiveness of the app score per group with SDs.

[[PNG File , 19 KB - jmir_v24i3e32130_app5.png](#)]

Multimedia Appendix 6

CONSORT-eHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 1216 KB - jmir_v24i3e32130_app6.pdf](#)]

References

1. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. *J Med Internet Res* 2016;18(11):e287 [[FREE Full text](#)] [doi: [10.2196/jmir.5692](https://doi.org/10.2196/jmir.5692)] [Medline: [27806926](https://pubmed.ncbi.nlm.nih.gov/27806926/)]
2. Rose T, Barker M, Maria Jacob C, 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;61(6):669-677. [doi: [10.1016/j.jadohealth.2017.05.024](https://doi.org/10.1016/j.jadohealth.2017.05.024)] [Medline: [28822682](https://pubmed.ncbi.nlm.nih.gov/28822682/)]
3. Cotie LM, Prince SA, Elliott CG, Ziss MC, McDonnell LA, Mullen KA, et al. The effectiveness of eHealth interventions on physical activity and measures of obesity among working-age women: a systematic review and meta-analysis. *Obes Rev* 2018;19(10):1340-1358. [doi: [10.1111/obr.12700](https://doi.org/10.1111/obr.12700)] [Medline: [30156044](https://pubmed.ncbi.nlm.nih.gov/30156044/)]
4. Jonkman NH, van Schooten KS, Maier AB, Pijnappels M. eHealth interventions to promote objectively measured physical activity in community-dwelling older people. *Maturitas* 2018;113:32-39. [doi: [10.1016/J.MATURITAS.2018.04.010](https://doi.org/10.1016/J.MATURITAS.2018.04.010)] [Medline: [29903646](https://pubmed.ncbi.nlm.nih.gov/29903646/)]
5. Duff OM, Walsh DM, Furlong BA, O'Connor NE, Moran KA, Woods CB. Behavior change techniques in physical activity eHealth interventions for people with cardiovascular disease: systematic review. *J Med Internet Res* 2017;19(8):e281 [[FREE Full text](#)] [doi: [10.2196/jmir.7782](https://doi.org/10.2196/jmir.7782)] [Medline: [28768610](https://pubmed.ncbi.nlm.nih.gov/28768610/)]
6. Haberlin C, O'Dwyer T, Mockler D, Moran J, O'Donnell DM, Broderick J. The use of eHealth to promote physical activity in cancer survivors: a systematic review. *Support Care Cancer* 2018;26(10):3323-3336. [doi: [10.1007/s00520-018-4305-z](https://doi.org/10.1007/s00520-018-4305-z)] [Medline: [29909476](https://pubmed.ncbi.nlm.nih.gov/29909476/)]
7. Oosterveen E, Tzelepis F, Ashton L, Hutchesson MJ. A systematic review of eHealth behavioral interventions targeting smoking, nutrition, alcohol, physical activity and/or obesity for young adults. *Prev Med* 2017;99:197-206. [doi: [10.1016/j.ypmed.2017.01.009](https://doi.org/10.1016/j.ypmed.2017.01.009)] [Medline: [28130046](https://pubmed.ncbi.nlm.nih.gov/28130046/)]
8. Mönninghoff A, Kramer JN, Hess AJ, Ismailova K, Teepe GW, Tudor Car L, et al. Long-term effectiveness of mHealth physical activity interventions: systematic review and meta-analysis of randomized controlled trials. *J Med Internet Res* 2021;23(4):e26699. [doi: [10.2196/26699](https://doi.org/10.2196/26699)] [Medline: [33811021](https://pubmed.ncbi.nlm.nih.gov/33811021/)]
9. Spagnolo PA, Colloca L, Heilig M. The role of expectation in the therapeutic outcomes of alcohol and drug addiction treatments. *Alcohol Alcohol* 2015;50(3):285-285. [doi: [10.1093/alcalc/agt015](https://doi.org/10.1093/alcalc/agt015)] [Medline: [25761920](https://pubmed.ncbi.nlm.nih.gov/25761920/)]
10. Rief W, Shedden-Mora MC, Laferton JA, Auer C, Petrie KJ, Salzmann S, et al. Preoperative optimization of patient expectations improves long-term outcome in heart surgery patients: results of the randomized controlled PSY-HEART trial. *BMC Med* 2017;15(1):4. [doi: [10.1186/s12916-016-0767-3](https://doi.org/10.1186/s12916-016-0767-3)] [Medline: [28069021](https://pubmed.ncbi.nlm.nih.gov/28069021/)]

11. Enck P, Bingel U, Schedlowski M, Rief W. The placebo response in medicine: minimize, maximize or personalize? *Nat Rev Drug Discov* 2013;12(3):191-204. [doi: [10.1038/nrd3923](https://doi.org/10.1038/nrd3923)] [Medline: [23449306](https://pubmed.ncbi.nlm.nih.gov/23449306/)]
12. Bartels DJ, van Laarhoven AI, Haverkamp EA, Wilder-Smith OH, Donders AR, van Middendorp H, et al. Role of conditioning and verbal suggestion in placebo and nocebo effects on itch. *PLoS One* 2014;9(3):e91727. [doi: [10.1371/journal.pone.0091727](https://doi.org/10.1371/journal.pone.0091727)] [Medline: [24646924](https://pubmed.ncbi.nlm.nih.gov/24646924/)]
13. Benedetti F, Pollo A, Lopiano L, Lanotte M, Vighetti S, Rainero I. Conscious expectation and unconscious conditioning in analgesic, motor, and hormonal placebo/nocebo responses. *J Neurosci* 2003;23(10):4315-4323. [doi: [10.1523/jneurosci.23-10-04315.2003](https://doi.org/10.1523/jneurosci.23-10-04315.2003)] [Medline: [12764120](https://pubmed.ncbi.nlm.nih.gov/12764120/)]
14. Mondloch MV, Cole DC, Frank JW. Does how you do depend on how you think you'll do? A systematic review of the evidence for a relation between patients' recovery expectations and health outcomes. *CMAJ* 2001;165(2):174-179. [Medline: [11501456](https://pubmed.ncbi.nlm.nih.gov/11501456/)]
15. Wolters F, Peerdeman KJ, Evers AW. Placebo and nocebo effects across symptoms: from pain to fatigue, dyspnea, nausea, and itch. *Front Psychiatry* 2019;10:470 [FREE Full text] [doi: [10.3389/fpsy.2019.00470](https://doi.org/10.3389/fpsy.2019.00470)] [Medline: [31312148](https://pubmed.ncbi.nlm.nih.gov/31312148/)]
16. Colloca L, Miller FG. How placebo responses are formed: a learning perspective. *Philos Trans R Soc Lond B Biol Sci* 2011;366(1572):1859-1869 [FREE Full text] [doi: [10.1098/rstb.2010.0398](https://doi.org/10.1098/rstb.2010.0398)] [Medline: [21576143](https://pubmed.ncbi.nlm.nih.gov/21576143/)]
17. Stalujanis E, Neufeld J, Glaus Stalder M, Belardi A, Tegethoff M, Meinschmidt G. Induction of efficacy expectancies in an ambulatory smartphone-based digital placebo mental health intervention: randomized controlled trial. *JMIR Mhealth Uhealth* 2021;9(2):e20329 [FREE Full text] [doi: [10.2196/20329](https://doi.org/10.2196/20329)] [Medline: [33594991](https://pubmed.ncbi.nlm.nih.gov/33594991/)]
18. Increasing the efficiency of the Mobile Step Coach with positive verbal suggestions. *Osf registries*. 2020. URL: <https://osf.io/cwjjes> [accessed 2022-02-07]
19. Pollo A, Carlino E, Benedetti F. The top-down influence of ergogenic placebos on muscle work and fatigue. *Eur J Neurosci* 2008;28(2):379-388. [doi: [10.1111/j.1460-9568.2008.06344.x](https://doi.org/10.1111/j.1460-9568.2008.06344.x)] [Medline: [18702709](https://pubmed.ncbi.nlm.nih.gov/18702709/)]
20. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007;39(2):175-191. [doi: [10.3758/bf03193146](https://doi.org/10.3758/bf03193146)] [Medline: [17695343](https://pubmed.ncbi.nlm.nih.gov/17695343/)]
21. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: a novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: overview and preliminary evaluation in the public health context. In: 2015 Wireless Telecommunications Symposium. Piscataway: IEEE; 2015 Presented at: WTS '15; April 15-17, 2015; New York p. 1-6. [doi: [10.1109/wts.2015.7117255](https://doi.org/10.1109/wts.2015.7117255)]
22. Kowatsch T, Volland D, Shih I, Rügger D, Künzler F, Barata F, et al. Design and evaluation of a mobile chat app for the open source behavioral health intervention platform MobileCoach. In: Proceedings of the 12th International Conference on Design Science Research in Information System and Technology. 2017 Presented at: DESRIST '17; May 30-June 1, 2017; Karlsruhe p. 485-489. [doi: [10.1007/978-3-319-59144-5_36](https://doi.org/10.1007/978-3-319-59144-5_36)]
23. Kowatsch T, Schachner T, Harperink S, Barata F, Dittler U, Xiao G, et al. Conversational agents as mediating social actors in chronic disease management involving health care professionals, patients, and family members: multisite single-arm feasibility study. *J Med Internet Res* 2021;23(2):e25060. [doi: [10.2196/25060](https://doi.org/10.2196/25060)] [Medline: [33484114](https://pubmed.ncbi.nlm.nih.gov/33484114/)]
24. Kramer JN, 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;54(7):518-528. [doi: [10.1093/abm/kaaa002](https://doi.org/10.1093/abm/kaaa002)] [Medline: [32182353](https://pubmed.ncbi.nlm.nih.gov/32182353/)]
25. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. *JMIR Mhealth Uhealth* 2020;8(4):e15806. [doi: [10.2196/15806](https://doi.org/10.2196/15806)] [Medline: [32242820](https://pubmed.ncbi.nlm.nih.gov/32242820/)]
26. Stieger M, Flückiger C, Rügger D, Kowatsch T, Roberts BW, Allemand M. Changing personality traits with the help of a digital personality change intervention. *Proc Natl Acad Sci U S A* 2021;118(8):e2017548118. [doi: [10.1073/pnas.2017548118](https://doi.org/10.1073/pnas.2017548118)] [Medline: [33558417](https://pubmed.ncbi.nlm.nih.gov/33558417/)]
27. Nass C, Steuer J, Tauber ER. Computers are social actors. In: Proceedings of the SIGCHI conference on human factors in computing systems. 1994 Presented at: CHI '94; April 24-28, 1994; Boston p. 72-78. [doi: [10.1145/191666.191703](https://doi.org/10.1145/191666.191703)]
28. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design features of embodied conversational agents in eHealth: a literature review. *Int J Hum-Comput Stud* 2020;138:102409. [doi: [10.1016/j.ijhcs.2020.102409](https://doi.org/10.1016/j.ijhcs.2020.102409)]
29. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. *J Couns Psychol* 1989;36(2):223-233. [doi: [10.1037/0022-0167.36.2.223](https://doi.org/10.1037/0022-0167.36.2.223)]
30. Bickmore T, Gruber A, Picard R. Establishing the computer-patient working alliance in automated health behavior change interventions. *Patient Educ Couns* 2005;59(1):21-30. [doi: [10.1016/j.pec.2004.09.008](https://doi.org/10.1016/j.pec.2004.09.008)] [Medline: [16198215](https://pubmed.ncbi.nlm.nih.gov/16198215/)]
31. Del Re AC, Flückiger C, Horvath AO, Wampold BE. Examining therapist effects in the alliance-outcome relationship: a multilevel meta-analysis. *J Consult Clin Psychol* 2021;89(5):371-378. [doi: [10.1037/ccp0000637](https://doi.org/10.1037/ccp0000637)] [Medline: [33829817](https://pubmed.ncbi.nlm.nih.gov/33829817/)]
32. Flückiger C, Del Re AC, Wampold BE, Horvath AO. The alliance in adult psychotherapy: a meta-analytic synthesis. *Psychotherapy (Chic)* 2018;55(4):316-340. [doi: [10.1037/pst0000172](https://doi.org/10.1037/pst0000172)] [Medline: [29792475](https://pubmed.ncbi.nlm.nih.gov/29792475/)]
33. Flückiger C, Del Re AC, Wampold BE, Symonds D, Horvath AO. How central is the alliance in psychotherapy? A multilevel longitudinal meta-analysis. *J Couns Psychol* 2012;59(1):10-17. [doi: [10.1037/a0025749](https://doi.org/10.1037/a0025749)] [Medline: [21988681](https://pubmed.ncbi.nlm.nih.gov/21988681/)]

34. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. *Am J Health Promot* 1997;12(1):38-48. [doi: [10.4278/0890-1171-12.1.38](https://doi.org/10.4278/0890-1171-12.1.38)] [Medline: [10170434](https://pubmed.ncbi.nlm.nih.gov/10170434/)]
35. Michie S, Johnston M. Behavior change techniques. In: Gellman MD, Turner RJ, editors. *Encyclopaedia of Behavioral Medicine*. New York: Springer; 2013:182-187.
36. Kelders SM, Kip H. Development and initial validation of a scale to measure engagement with eHealth technologies. In: *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019 Presented at: CHI EA '19; May 4-9, 2019; Glasgow p. 1-6. [doi: [10.1145/3290607.3312917](https://doi.org/10.1145/3290607.3312917)]
37. Ryan RM, Frederick C. On energy, personality, and health: subjective vitality as a dynamic reflection of well-being. *J Pers* 1997;65(3):529-565. [doi: [10.1111/j.1467-6494.1997.tb00326.x](https://doi.org/10.1111/j.1467-6494.1997.tb00326.x)] [Medline: [9327588](https://pubmed.ncbi.nlm.nih.gov/9327588/)]
38. Vercoulen JH, Alberts M, Bleijenberg G. De checklist individuele spankracht (CIS). *Gedragstherapie* 1999;32(131):6 [FREE Full text]
39. Markland D, Tobin V. A modification to the behavioural regulation in exercise questionnaire to include an assessment of amotivation. *J Sport Exerc Psychol* 2004;26(2):191-196. [doi: [10.1123/jsep.26.2.191](https://doi.org/10.1123/jsep.26.2.191)]
40. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med* 2006;166(10):1092-1097. [doi: [10.1001/archinte.166.10.1092](https://doi.org/10.1001/archinte.166.10.1092)] [Medline: [16717171](https://pubmed.ncbi.nlm.nih.gov/16717171/)]
41. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd Edition. Milton Park: Routledge; 1988.
42. Marcus BH, Banspach SW, Lefebvre RC, Rossi JS, Carleton RA, Abrams DB. Using the stages of change model to increase the adoption of physical activity among community participants. *Am J Health Promot* 1992;6(6):424-429. [doi: [10.4278/0890-1171-6.6.424](https://doi.org/10.4278/0890-1171-6.6.424)] [Medline: [10146803](https://pubmed.ncbi.nlm.nih.gov/10146803/)]
43. Calfas KJ, Sallis JF, Oldenburg B, Ffrench M. Mediators of change in physical activity following an intervention in primary care: PACE. *Prev Med* 1997;26(3):297-304. [doi: [10.1006/pmed.1997.0141](https://doi.org/10.1006/pmed.1997.0141)] [Medline: [9144753](https://pubmed.ncbi.nlm.nih.gov/9144753/)]
44. Steptoe A, Kerry S, Rink E, Hilton S. The impact of behavioral counseling on stage of change in fat intake, physical activity, and cigarette smoking in adults at increased risk of coronary heart disease. *Am J Public Health* 2001;91(2):265-269. [doi: [10.2105/ajph.91.2.265](https://doi.org/10.2105/ajph.91.2.265)] [Medline: [11211636](https://pubmed.ncbi.nlm.nih.gov/11211636/)]
45. Peerdeman KJ, van Laarhoven AI, Peters ML, Evers AW. An integrative review of the influence of expectancies on pain. *Front Psychol* 2016;7:1270 [FREE Full text] [doi: [10.3389/fpsyg.2016.01270](https://doi.org/10.3389/fpsyg.2016.01270)] [Medline: [27602013](https://pubmed.ncbi.nlm.nih.gov/27602013/)]
46. Vase L, Wartolowska K. Pain, placebo, and test of treatment efficacy: a narrative review. *Br J Anaesth* 2019;123(2):e254-e262 [FREE Full text] [doi: [10.1016/j.bja.2019.01.040](https://doi.org/10.1016/j.bja.2019.01.040)] [Medline: [30915982](https://pubmed.ncbi.nlm.nih.gov/30915982/)]
47. Meeuwis SH, van Middendorp H, Veldhuijzen DS, van Laarhoven AI, De Houwer J, Lavrijsen AP, et al. Placebo effects of open-label verbal suggestions on itch. *Acta Derm Venereol* 2018;98(2):268-274. [doi: [10.2340/00015555-2823](https://doi.org/10.2340/00015555-2823)] [Medline: [29057429](https://pubmed.ncbi.nlm.nih.gov/29057429/)]
48. Meeuwis SH, van Middendorp H, van Laarhoven AI, Veldhuijzen DS, Lavrijsen AP, Evers AW. Effects of open- and closed-label nocebo and placebo suggestions on itch and itch expectations. *Front Psychiatry* 2019;10:436 [FREE Full text] [doi: [10.3389/fpsyg.2019.00436](https://doi.org/10.3389/fpsyg.2019.00436)] [Medline: [31293458](https://pubmed.ncbi.nlm.nih.gov/31293458/)]
49. Müller V, Remus K, Hoffmann V, Tschöp MH, Meissner K. Effectiveness of a placebo intervention on visually induced nausea in women – a randomized controlled pilot study. *J Psychosom Res* 2016;91:9-11. [doi: [10.1016/j.jpsychores.2016.09.012](https://doi.org/10.1016/j.jpsychores.2016.09.012)] [Medline: [27894470](https://pubmed.ncbi.nlm.nih.gov/27894470/)]
50. Weimer K, Schulte J, Maichle A, Muth ER, Scisco JL, Horing B, et al. Effects of ginger and expectations on symptoms of nausea in a balanced placebo design. *PLoS One* 2012;7(11):e49031. [doi: [10.1371/journal.pone.0049031](https://doi.org/10.1371/journal.pone.0049031)] [Medline: [23152846](https://pubmed.ncbi.nlm.nih.gov/23152846/)]
51. Isenberg SA, Lehrer PM, Hochron S. The effects of suggestion on airways of asthmatic subjects breathing room air as a suggested bronchoconstrictor and bronchodilator. *J Psychosom Res* 1992;36(8):769-776. [doi: [10.1016/0022-3999\(92\)90135-o](https://doi.org/10.1016/0022-3999(92)90135-o)] [Medline: [1432867](https://pubmed.ncbi.nlm.nih.gov/1432867/)]
52. Bingel U. Placebo 2.0: the impact of expectations on analgesic treatment outcome. *Pain* 2020;161 Suppl 1:S48-S56. [doi: [10.1097/j.pain.0000000000001981](https://doi.org/10.1097/j.pain.0000000000001981)] [Medline: [33090739](https://pubmed.ncbi.nlm.nih.gov/33090739/)]
53. Darragh M, Chang JW, Booth RJ, Considine NS. The placebo effect in inflammatory skin reactions: the influence of verbal suggestion on itch and weal size. *J Psychosom Res* 2015;78(5):489-494. [doi: [10.1016/j.jpsychores.2015.01.011](https://doi.org/10.1016/j.jpsychores.2015.01.011)] [Medline: [25649275](https://pubmed.ncbi.nlm.nih.gov/25649275/)]
54. Quinn VF, Colagiuri B. Placebo interventions for nausea: a systematic review. *Ann Behav Med* 2015;49(3):449-462. [doi: [10.1007/s12160-014-9670-3](https://doi.org/10.1007/s12160-014-9670-3)] [Medline: [25515086](https://pubmed.ncbi.nlm.nih.gov/25515086/)]
55. Sher L. The placebo effect on mood and behavior: the role of the endogenous opioid system. *Med Hypotheses* 1997;48(4):347-349. [doi: [10.1016/s0306-9877\(97\)90105-5](https://doi.org/10.1016/s0306-9877(97)90105-5)] [Medline: [9160290](https://pubmed.ncbi.nlm.nih.gov/9160290/)]
56. Kirsch I. Response expectancy and the placebo effect. *Int Rev Neurobiol* 2018;138:81-93. [doi: [10.1016/bs.irm.2018.01.003](https://doi.org/10.1016/bs.irm.2018.01.003)] [Medline: [29681336](https://pubmed.ncbi.nlm.nih.gov/29681336/)]
57. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. *Psychol Rev* 1977;84(2):191-215. [doi: [10.1037/0033-295X.84.2.191](https://doi.org/10.1037/0033-295X.84.2.191)] [Medline: [847061](https://pubmed.ncbi.nlm.nih.gov/847061/)]
58. Bandura A, Freeman WH, Lightsey R. Self-efficacy: the exercise of control. *J Cogn Psychother* 1999;13(2):158-166. [doi: [10.1891/0889-8391.13.2.158](https://doi.org/10.1891/0889-8391.13.2.158)]

59. Schakel L, Veldhuijzen DS, van Middendorp V, Van Dessel P, De Houwer J, Bidarra R, et al. The effects of a gamified approach avoidance training and verbal suggestions on food outcomes. *PLoS One* 2018;13(7):e0201309. [doi: [10.1371/journal.pone.0201309](https://doi.org/10.1371/journal.pone.0201309)] [Medline: [30048511](https://pubmed.ncbi.nlm.nih.gov/30048511/)]
60. Schakel L, Veldhuijzen DS, Manai M, van Beugen S, van der Vaart R, van Middendorp H, et al. Editor's choice: optimizing healthy food preferences by serious gaming. *Psychol Health* 2020;35(4):405-424. [doi: [10.1080/08870446.2019.1675657](https://doi.org/10.1080/08870446.2019.1675657)] [Medline: [31607172](https://pubmed.ncbi.nlm.nih.gov/31607172/)]
61. Fiorio M, Emadi Andani M, Marotta A, Classen J, Tinazzi M. Placebo-induced changes in excitatory and inhibitory corticospinal circuits during motor performance. *J Neurosci* 2014;34(11):3993-4005. [doi: [10.1523/jneurosci.3931-13.2014](https://doi.org/10.1523/jneurosci.3931-13.2014)] [Medline: [24623777](https://pubmed.ncbi.nlm.nih.gov/24623777/)]
62. Schuetzler RM, Giboney JS, Grimes GM, Nunamaker Jr JF. The influence of conversational agent embodiment and conversational relevance on socially desirable responding. *Decis Support Syst* 2018;114:94-102. [doi: [10.1016/j.dss.2018.08.011](https://doi.org/10.1016/j.dss.2018.08.011)]
63. Pickard MD, Roster CA. Using computer automated systems to conduct personal interviews: does the mere presence of a human face inhibit disclosure? *Comput Hum Behav* 2020;105:106197. [doi: [10.1016/j.chb.2019.106197](https://doi.org/10.1016/j.chb.2019.106197)]

Edited by A Mavragani; submitted 15.07.21; peer-reviewed by J Trevino, M Barroso Hurtado, M Khokhar; comments to author 18.08.21; revised version received 29.09.21; accepted 07.12.21; published 01.03.22.

Please cite as:

Skvortsova A, Cohen Rodrigues T, de Buissonjé D, Kowatsch T, Santhanam P, Veldhuijzen DS, van Middendorp H, Evers A
Increasing the Effectiveness of a Physical Activity Smartphone Intervention With Positive Suggestions: Randomized Controlled Trial
J Med Internet Res 2022;24(3):e32130

URL: <https://www.jmir.org/2022/3/e32130>

doi: [10.2196/32130](https://doi.org/10.2196/32130)

PMID: [35230245](https://pubmed.ncbi.nlm.nih.gov/35230245/)

©Aleksandrina Skvortsova, Talia Cohen Rodrigues, David de Buissonjé, Tobias Kowatsch, Prabhakaran Santhanam, Dieuwke S Veldhuijzen, Henriët van Middendorp, Andrea Evers. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 01.03.2022. 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.

Original Paper

Effect of Source Type and Protective Message on the Critical Evaluation of News Messages on Facebook: Randomized Controlled Trial in the Netherlands

Frans Folkvord¹, PhD; Freek Snelting², MSc; Doeschka Anschutz³, PhD; Tilo Hartmann⁴, PhD; Alexandra Theben⁵, PhD; Laura Gunderson¹, MSc; Ivar Vermeulen⁴, PhD; Francisco Lupiáñez-Villanueva¹, PhD

¹Open Evidence Research, Open Evidence, Barcelona, Spain

²Nijmegen, Netherlands

³Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

⁴Communication Science, Vrije Universiteit, Amsterdam, Netherlands

⁵Barcelona school of management, UPF Barcelona School of Management, Barcelona, Spain

Corresponding Author:

Frans Folkvord, PhD

Open Evidence Research

Open Evidence

Avinguda Tibidabo 35

Barcelona, 08035

Spain

Phone: 34 31610948122

Email: fransfolkvord@gmail.com

Abstract

Background: Disinformation has become an increasing societal concern, especially due to the speed that news is shared in the digital era. In particular, disinformation in the health care sector can lead to serious casualties, as the current COVID-19 crisis clearly shows.

Objective: The main aim of this study was to experimentally examine the effects of information about the source and a protective warning message on users' critical evaluation of news items, as well as the perception of accuracy of the news item.

Methods: A 3 (unreliable vs reliable vs no identified source) × 2 (with protective message vs without) between-subject design was conducted among 307 participants (mean age 29 (SD 10.9) years).

Results: The results showed a significant effect of source information on critical evaluation. In addition, including a protective message did not significantly affect critical evaluation. The results showed no interaction between type of source and protective message on critical evaluation.

Conclusions: Based on these results, it is questionable whether including protective messages to improve critical evaluation is a way to move forward and improve critical evaluation of health-related news items, although effective methodologies to tackle the spread of disinformation are highly needed.

Trial Registration: ClinicalTrials.gov NCT05030883; <https://clinicaltrials.gov/ct2/show/NCT05030883>

(*J Med Internet Res* 2022;24(3):e27945) doi:[10.2196/27945](https://doi.org/10.2196/27945)

KEYWORDS

health communication; disinformation; protective message; source; critical evaluation; critical thinking

Introduction

Background

Disinformation is an increasing societal concern in many countries and had been highly apparent during the COVID-19

crisis [1-5]. Following the definition established by UNESCO (United Nations Educational, Scientific and Cultural Organization) [4], disinformation is considered in this study as information that is false and deliberately created to harm a person, social group organization, or country [4-6], while misinformation is defined as information that is false but has

not been created with the intention to harm someone and misinformation is defined as information based on reality and used to inflict harm on a person, social group, organization, or country. During times of crises, such as the current COVID-19 situation, disinformation in the health space is thriving. Therefore, European institutions are jointly fighting against disinformation and working in close cooperation with online platforms to ensure effective public health communication [1,7,8].

Transparency regarding the way information is produced or sponsored is necessary to support high quality (digital) information provision, and the relation between information creators and distributors needs some rebalancing [9]. Disinformation may lead to inaccurate beliefs and can exacerbate partisan disagreement over basic facts that could easily be checked [10]. There may have been instances of dissemination of false or misleading information in the past, but in contemporary societies its consequences arrive faster, become more visible, and therefore seem to harm current societies to a larger extent than before. In particular, Facebook was used as a platform to direct users to websites where disinformation was published, showing the importance of the hierarchical, institutional, or professional structures that traditional media provides as a gatekeeper but social media platforms miss because of their origin [11,12].

Literature Review

Following the truth-default theory [13], people generally tend to believe others as a default state, a so-called “truth bias.” Levine [13] proposes that this truth default is adaptive, based on contextual and informational factors. Considering that belief in honesty is a default state, which is a passive starting place for making inferences about communication, the truth-default theory predicts that once suspicion is actively triggered, the possibility of processing a message as deception might occur.

Nowadays, social media platforms have become de facto news curators [14,15]. Since many people, particularly youth [3], consume their main news information from social media platforms, the influence of disinformation has become a real threat around the world and should be investigated [16-19]. For example, about half of American adults (53%) report that they get news from social media often or sometimes [17].

Due to the information flows provided by the internet, information today can be published by anyone and come from anywhere, which is a participatory strength of an open society but also has its weaknesses [20,21]. Disinformation often contains appealing titles and salient pictures that cue emotional arousal, giving rise to impulsive sharing decisions [22]. Consequently, disinformation is often quickly shared on social media after the sharer has read the attractive title or appealing highlights but perhaps not the complete content [23].

Empirical Evidence

Regardless of why the information was shared, this phenomenon only reinforces the problem of the dissemination of disinformation [20]. Vosoughi et al [24] demonstrate that the rate of spreading of disinformation is 6 times that of spreading truthful content. Bode and Vraga [25,26] suggest that once

disinformation is absorbed, it is much more difficult to change the misperception or belief (ie, rectify it) and that it can take a long time to resolve false rumors [27]. In particular, youths find it more difficult to overcome the spread of disinformation, as they are digital natives and most fake news is shared on online platforms [28]. Another study showed university students visiting social media platforms more often, and sharing and liking posts increases the susceptibility of students to fake news [29]. Furthermore, Buchanan showed that sharing false information online was not influenced by authoritativeness of the source of the material, and participants' level of digital literacy had little effect on their responses [30]. Importantly, multimodal disinformation, which occurs mainly on social media, is considered slightly more credible than textual disinformation [31].

A meta-analysis by Chan et al [32] shows that media literacy can affect the extent to which people can interpret disinformation. The researchers recommend improving people's media literacy to encourage a critical attitude toward the processing of disinformation. Media literacy allows users to be more critical of what they see and read, so they can judge for themselves whether they believe the information [32]. More specifically, eHealth literacy, the competence of people to accurately comprehend health information, also influences individual susceptibility to health-related information provided on social media [33].

In addition, recent research by Clayton et al [34] shows that disinformation on Facebook is perceived to be less accurate when a protective warning message is posted announcing that disinformation occurs on this platform. The authors argue that more research needs to be done to ensure disinformation is effectively combated to reduce the negative effects on health care provision, but that real and accurate news does not suffer from the effects of such an intervention [34]. Therefore, it seems important to focus on stimulating critical processing of the content of news messages rather than just using a warning that influences the perception of accuracy of news messages in general. Unfortunately, relatively little is known about effective interventions that cause Facebook users to evaluate information more critically, thus enabling them to recognize disinformation more easily.

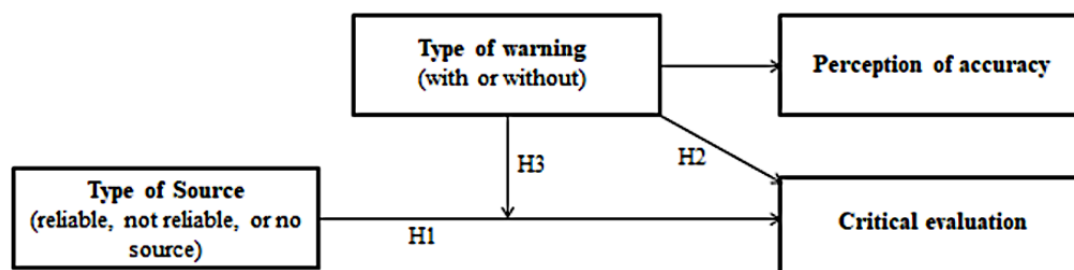
Study Objective and Hypotheses

The main objective of this study is to gain more insight into the extent to which source information and a protective message have an effect on critical evaluation by Dutch Facebook users. In [Figure 1](#) we show the conceptual model. Based on the truth-default theory, the expectation is that participants exposed to news messages from an unreliable source will activate more critical reflection than participants exposed to news messages from a reliable or unidentified source (H1a). Furthermore, it is hypothesized that participants exposed to news messages from an unreliable source will perceive these news items as less accurate than participants exposed to news messages from a reliable or unidentified source (H1b). In addition, participants exposed to news messages following a protective message will activate more critical reflection (H2a) and perceive these items as less accurate (H2b) than participants exposed to news

messages without a preceding protective message. Finally, it is expected that adding a protective message will moderate the effect of the credibility of the source on critical reflection and perceived accuracy of news items. More specifically, we expect that participants exposed to news messages from an unreliable source who have previously encountered a protective message will activate more critical reflection (H3a) and perceive news

items as less accurate (H3b) than participants who have not encountered a protective message before. In contrast, we do not expect any difference in critical reflection and perceived accuracy to be instigated by the protective message among participants exposed to news messages from a reliable or unidentified source.

Figure 1. Conceptual model.



Methods

Design

To test the hypotheses and investigate the effects of source information and a protective message on critical evaluation and perception of accuracy of news items on Facebook, an online between-subjects experiment with 3 (unreliable vs reliable vs unidentified source) \times 2 (with protective message vs without) levels was constructed. Six conditions were created in line with the factorial design of the study, and participants were randomly assigned to one of the conditions. Participants' critical evaluation and the perception of accuracy of the presented (news) messages were assessed as dependent variables.

Stimulus Material

The stimulus material consisted of a short video clip showing how an (anonymous) user opens the Facebook app on a smartphone to scroll over the Facebook timeline. Considering that many people consume their main news from social media [17], we decided to use Facebook as stimulus material. This is a daily activity for many Facebook users, and the video clip aimed to imitate this daily activity as realistically as possible. The Facebook timeline featured in the clip contained 3 short (news) messages shown one after the other. Each (news) message was on screen for 10 seconds. No other messages were shown in the video.

The (news) messages were about ambiguous health-related topics so it would not be immediately apparent to participants whether the news messages was accurate [11]. This setup made it possible to test the effects of the source type and protective message. The presented (news) messages addressed vaccination [35], climate change [36], and health insurance [37]. The (news) messages about vaccination and health insurance were made up for this study, and the climate report was a real news report from a Dutch public broadcaster, Nederlands Omroep Stichting (NOS) [38]. The headline for the vaccination item was "Prohibition for unvaccinated children" and the body text was "Is your child not vaccinated? Then he or she is no longer welcome at daycare centers and primary schools. This decision

made last Monday. There is an unnecessarily high risk and that is why this ban is necessary, the government says." The headline for the climate change item was "Dutch aviation's impact on climate change is going to be enormous" and the body text was "If no new measures are taken, the CO₂ emissions of Dutch aviation will double over the next 30 years. This will make aviation one of the largest polluters in the Netherlands in 2050." The headline for the health insurance item was "Deductible amount in health care will double next year" and the body text was "Health insurers can no longer finance high health care costs if people's deductible amount remains so low. Deductible amount must increase considerably in order to be able to provide good care for everyone."

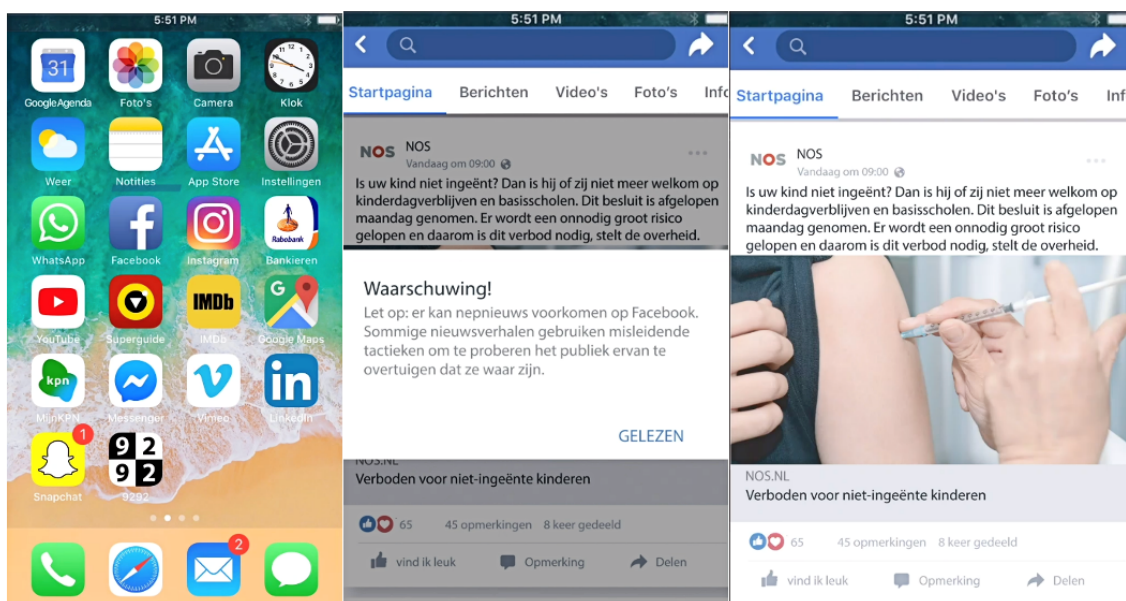
Source Information

To examine the influence of source information, both above and below each news message an unreliable, reliable, or unidentified source was provided. The choice of the unreliable source was based on previous research by Pennycook and Rand [39], who used the existing news site Dailybuzzlive to represent an unreliable source. The reliable source chosen was NOS, based on research by the Media Monitor [40] showing that NOS is perceived as the most reliable news source in the Netherlands. Unidentified source (news) messages were shown without source information.

Protective Message

To examine the influence of a protective message, half of the participants received a protective message in the video displaying the (news) messages. After the anonymous users displayed in the video opened the Facebook app, a warning about fake news appeared. This warning was in Dutch and motivated by previous research by Clayton et al [34]. It read "Warning! Note: fake news can occur on Facebook. Some news stories use misleading tactics to try and convince the public that they are true." The warning was on the screen for 10 seconds, so the short video clip with a warning lasted 10 seconds longer (48 seconds in total) than the clip without a warning. Figure 2 shows an example of how a reliable source with a protective message was presented.

Figure 2. Stimulus material.

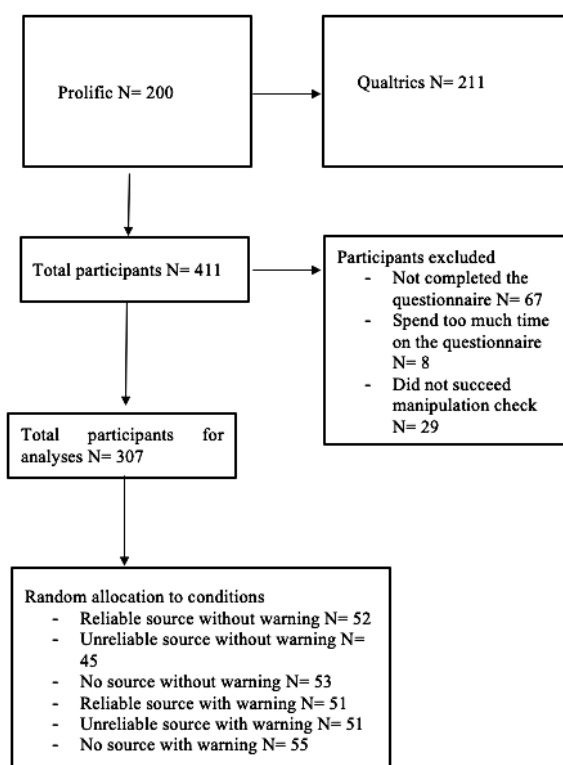


Participants

Participants were recruited during May and June 2019 through an online panel (Prolific) and a Qualtrics link shared on social media channels such as Facebook, Instagram, and LinkedIn. Analyses were conducted separately for both groups of participants, but no different effects were found. Therefore, we have decided to aggregate the two groups and only report the analyses for the complete group of participants. A G*power analysis (effect size=0.25, alpha=.05, power=.80, and two predictors) determined a minimum sample size of 158 participants [41]. In total, 411 participants were recruited to the study, however, 67 participants did not fully complete the

questionnaire. In the experiment, it was important that participants completed the questionnaire without pausing because this could interrupt the flow and would harm the validation of the protective message. In order for the effect of the manipulation to be tested accurately, participants (n=8) who needed more than 15 minutes to complete the questionnaire were removed. The control question was a check to see if participants paid attention to the stimulus material, and participants (n=29) who could not remember what the news items were about (assessed with a closed question) were removed. After dropping participants who did not meet the abovementioned standards, the total number of participants was 307 (see Figure 3).

Figure 3. Flow diagram of study participants.



Ethics Approval

The ethical committee of the Tilburg School of Humanities and Digital Sciences (REDC 2019-60, obtained April 29, 2019) in the Netherlands approved this study. This trial was registered at ClinicalTrials.gov [NCT05030883].

Procedure

Participants conducted the experiment in an online survey in Prolific and Qualtrics that included a random allocation procedure. Furthermore, the participants participated anonymously, and informed consent was provided by the participants. Next, participants were asked to report whether they had a Facebook account, and sociodemographic data were collected (gender, age, education level). Next, participants were randomly allocated to one of the 6 experimental conditions. After the video clip, users' general critical evaluation of the 3 displayed (news) messages was assessed. This was followed by questions about the perception of accuracy. Additional questions were then asked, such as reasons for Facebook use, frequency of Facebook use, whether participants follow the news via Facebook, whether they trust the news on Facebook, and to what extent they generally follow the news. Finally, a manipulation check was performed. People were asked if participants had seen a warning and what this warning was about and if they had observed a news source and what news source it was. After the last question a debriefing explained the purpose

of the research. After the debriefing, participants were thanked for their participation.

Measurements

Dependent Variables

Critical Evaluation Skills

Critical evaluation was assessed based on a scale developed by Metzger et al [41]. This unidimensional scale consists of 9 items on 5 different aspects (accuracy, authority, objectivity, currency, and coverage), testing on a 5-point Likert scale (1=never, 5=always) if participants conducted these behaviors when reading news (Table 1). To confirm the unidimensional structure reported by Metzger et al [41], a principal component analysis (PCA) with oblimin rotation was performed across the 9 variables. We decided to use PCA over confirmatory factor analysis because we were interested in all the variances of the data we collected. The Kaiser-Meyer-Olkin test showed that it was an adequate sample to perform a PCA, sample adequacy is .91 (great according to Field [41]), Bartlett test of sphericity was significant ($\chi^2_{36}=1573.62, P<.001$), and all variable loadings were $>.606$. The analysis yielded a single factor (Eigenvalue 5.15) that explained 57.28% of the total variance. Factor loadings are shown in Table 1. Hence, all 9 items were collapsed into a mean index, Cronbach $\alpha=.90$, mean 3.15 [SD .93].

Table 1. Principal component analysis critical evaluation.

Item	Component loading
Check if the information is up to date	0.730
Check the source	0.820
Verify the quality of the source	0.871
Check the origin of the source	0.818
Find out if the source is officially recognized	0.774
Consider whether the information presented is fact or opinion	0.739
Find out or try to understand the intent of the author	0.606
Find other sources to validate the information	0.625
Verify that the information is complete and comprehensive	0.786

Perception of Accuracy

This variable is based on earlier research by Clayton et al [34] and consisted of 1 question per (news) message. The aim of the question was to find out what perception of accuracy participants had with a (news) message. The participant was asked to what extent they thought the statements in the (news) message were accurate. This was measured on a 4-point Likert scale (1=not at all accurate, 4=very accurate).

Control Variables

Multiple variables were assessed as potential covariates. For example, participants were asked how often they use Facebook (1=daily, 7=never; based on Clayton et al [34]), whether they followed the news via Facebook (1=always, 5=never; based on the Pew Research Center [3]), and if they had confidence in news on Facebook (1=completely disagree, 5=completely agree)

[41]. In addition, a second statement asked whether participants followed the news on a daily basis (1=completely disagree, 5=completely agree) [3]. Finally, participants were asked their gender (male/female), age (numerical), and the highest level of education completed.

Manipulation Check

Participants answered 4 questions at the end of the experiment to check whether manipulation was observed. First, they were asked if they had seen a warning in the video. Participants could answer this question with yes or no. When participants had seen a warning, they were then asked what the warning was about. There were 4 choices: internet limit, fake news, virus, and changes in news. Second, participants were asked whether they had seen a news source in the video. Here too, participants were able to answer yes or no. When participants had seen a news source, they were asked which news source they had seen. They

could choose from Volkskrant, Dailybuzzlive, NOS, NederlandsNieuws, and I don't remember.

Results

Descriptive Information

Study participants were 53.7% (165/307) men and 46.3% (142/307) women. The average age of the participants was 29 [SD 10.90] years. The highest completed level of education of participants was mainly higher educated, with 0.3% (9/307) reporting primary school, primary education, or no education as highest educational level; 1.3% (40/307) reporting preparatory secondary vocational education or lower general secondary education; 15.0% (46/307) reporting school of higher general secondary education or the preuniversity education; 7.5% (23/307) reporting secondary vocational education; 37.8% (116/307) reporting higher professional education; and 38.1% (117/307) reporting university. All participants had a Facebook account. Participants mainly used Facebook to stay in touch with friends and family (221/307, 72.0%) and to stay informed about events (148/307, 48.2%). In addition, 25.7% (79/307) of participants indicated that they used Facebook for news and information, 20.5% (63/307) for pastimes, and 12.4% (38/307) to be able to follow brands or famous people. On average, participants spend a few times a week on Facebook (mean 2.08 [SD 1.50]) and on average they occasionally followed the news via Facebook (mean 3.74 [SD 1.03]). In addition, confidence in news via Facebook was low (mean 2.28 [SD .91]) and participants on average followed the news daily (mean 3.68 [SD 1.16]).

Correlational Analyses

To test for potential covariates, Pearson correlations were calculated between all control variables and critical evaluation and perceived accuracy of the news item. Only "trust in news via Facebook" ($r=-.140$, $P=.01$) and "the daily following of news" ($r=.208$, $P<.001$) were significantly related to critical evaluation. Participants with lower trust in Facebook news and greater daily news consumption engaged in more critical evaluation of the presented (news) messages. Therefore, "trust in news via Facebook" and "daily monitoring of news" were included as covariates in the analyses.

Manipulation Checks

First, a manipulation check on the source information showed that 88.3% (91/103) of participants in the conditions with a reliable source saw a news source, while 11.7% (12/103) did not. In conditions with an unreliable source, 63% (60/96) of participants indicated that they had seen a news source, and 38% (36/96) did not. In the conditions without a news source, 64.8% (69/108) indicated that they had not seen a news source and 36% (39/108) indicated that they had seen a news source. The chi-square test ($\chi^2_2=61.02$, $P<.001$) showed that the differences between the experimental conditions, whether or not participants saw a news source, were significant. This means that the manipulation, whether participants have seen a news source or not, has been successful. Nonetheless, we tested the hypotheses with and without the noncompliers. We found no

different in the results with and without noncompliers. Therefore, we decided to include all in the final analysis.

Second, of the participants who have seen a reliable news source, 96.7% (88/91) named NOS.nl as a source. In the unreliable source condition, 95% (57/60) indicated that they had seen Dailybuzzlive.nl. A salient feature is that 36.1% (39/108) of participants in the no source condition indicated that they had seen a certain news source, but this was not the case. For example, 7 participants indicated that they had seen Volkskrant.nl, 6 participants Dailybuzzlive.nl, 15 participants NOS.nl, 5 participants NederlandsNieuws.nl, and 6 participants could not remember.

Third, when asked whether participants had seen a warning in the video on Facebook, in the conditions without a warning, 92.7% (139/150) of participants indicated that they had not seen a warning. The remaining participants (11/150, 7.3%) indicated that they had seen a warning. In the conditions with a protective message, 91.1% (143/157) of participants indicated that they had seen a warning and 8.9% (14/157) indicated that they had not seen a warning. The chi-square test ($\chi^2_1=215.22$, $P<.001$) showed that the differences between the experimental conditions, whether participants had seen a warning or not, were significant. From this it can be concluded that the manipulation was successful. Of the participants who saw a protective message, 97.2% (139/143) indicated that the warning was about fake news.

Testing of the Hypotheses

An analysis of covariance (ANCOVA) on critical evaluation showed a significant main effect of source information on critical evaluation ($F_{2,299}=6.401$, $P=.02$). Post hoc analysis with Bonferroni correction showed that participants exposed to a news message with an unreliable source (mean 3.31, SE .09) scored significantly higher ($P=.02$) on critical evaluation than the group with a reliable source (mean 2.95, SE .09). There was no significant difference between no source (mean 3.19, SE .09) and a reliable ($P=.18$) or unreliable source ($P>.99$). These results partly support H1a. Furthermore, the ANCOVA showed that there is a marginally significant relation between trust in news via Facebook and critical evaluation ($F_{1,299}=3.62$, $P=.06$) and a significant relation between following the news daily and critical evaluation ($F_{1,299}=15.14$, $P<.001$).

In addition, the ANCOVA showed that a protective message did not significantly affect critical evaluation ($F_{1,299}=1.59$, $P=.21$), thereby rejecting H2a. Finally, the ANCOVA showed that there was no interaction effect of type of source and protective message on critical evaluation ($F_{2,299}=0.41$, $P=.67$), thereby rejecting H3a.

Next, we tested whether a protective message had an influence on the perception of accuracy in (news) messages on Facebook. A multivariate analysis of covariance (MANCOVA) analysis showed that there was no significant difference between groups that had seen a protective message and groups that had not on the perception of accuracy (Wilks $\Lambda=0.998$, $F_{3,303}=0.24$, $P=.87$), thereby rejecting H1b. Univariate tests showed that these groups did not differ on the (news) message about vaccination

($F_{1,305}=0.08$, $P=.77$, $\eta^2<.001$), climate change ($F_{1,305}=0.72$, $P=.40$, $\eta^2=.002$), and health insurance ($F_{1,305}=0.07$, $P=.79$, $\eta^2<.001$). This means that a protective message has no effect on the perception of accuracy.

Further, the MANCOVA analysis showed that there was a significant difference between source information on the perception of accuracy (Wilks $\Lambda=0.929$, $F_{2,304}=3.79$, $P<.001$). The outcomes of univariate tests show that the perception of accuracy by source information in the (news) message differed significantly for vaccination ($F_{2,304}=8.02$, $P<.001$, $\eta^2=.050$) and health insurance ($F_{2,304}=4.01$, $P=.02$, $\eta^2=.026$). There was no significant effect on the (news) message about climate ($F_{2,304}=2.34$, $P=.10$, $\eta^2=.015$). These partly confirm H2b.

To see which conditions of source information differed significantly from each other on perception of accuracy, a post hoc analysis with Bonferroni correction was performed. The results showed that there was a significant difference between a reliable and an unreliable source. The post hoc analysis showed that the group with a reliable source (mean 2.85, SE .08) scored significantly higher in the (news) message about vaccination ($P<.001$) on perception of accuracy than the group with an unreliable source (mean 2.39, SE .09). There was a marginally significant difference between no source (mean 2.58, SE .08) and a reliable source ($P=.06$) and no significant difference between no source and an unreliable source ($P=.27$). There was also an effect on the (news) message about health insurance. The group with a reliable source (mean 2.33, SE .08) scored significantly higher ($P=.04$) on perception of accuracy than the group with an unreliable source (mean 2.03, SE .08). There was a marginally significant difference between no source (mean 2.06, SE .08) and a reliable source ($P=.05$) and no significant difference between no source and an unreliable source ($P>.99$). This means that source information has an effect on the perception of accuracy in the (news) messages about vaccination and health insurance.

Finally, the MANCOVA showed that there was no interaction effect on type of source and protective message on perceived accuracy of all 3 messages ($P>.05$), thereby rejecting H3b.

Discussion

Principal Findings

The main objective of this study was to gain insight into the extent to which source information and a protective message have an effect on the critical evaluation by Dutch Facebook users. This study showed that source information has an effect on the extent to which someone critically evaluates (news) messages on Facebook. In accordance with the first hypothesis, participants more critically evaluated a (news) message when exposed to an unreliable source compared to a reliable source. No significant differences were found between an unreliable or reliable source and an unidentified source. In accordance with Metzger et al [41], the findings of this study also imply that participants use cognitive heuristics to assess information by critically evaluating to a greater or lesser extent. Source

credibility as a heuristic seems to play a determining role in the extent to which someone tends to critically evaluate a (news) message on Facebook [42,43].

In addition, we expected that a protective message preceding disinformation on Facebook would have a positive effect on the critical evaluation of participants. The truth-default theory [13] predicted that once suspicion was actively triggered, the possibility of processing a message as deception might come to mind. No evidence for such an effect was found, which is why hypothesis 2 was rejected. In addition, no evidence was found that a protective message moderated the effect of source information on critical evaluation. This, therefore, rejected hypothesis 3. Current solutions provided by the European Commission and large social media platforms promising to include protective messages to improve critical evaluation to tackle the spread of disinformation may therefore be of only limited effectiveness, since most people believed the message was still valid and credible. This study showed that the inclusion of protective measures of this type might, in fact, not affect critical evaluation.

Despite positive results of a protective message found by others [2,34], we did not find any effect of a protective message on critically processing news information. A possible explanation for this is that participants' flow while reading the Facebook news had not been interrupted and, therefore, they had not processed the protective message with sufficient attention [44]. In addition, it is possible that people were not motivated or involved enough to critically evaluate the news messages on Facebook [45]. A suggestion for a follow-up study is that participants are obligated to look at the protective message and need to click on the protective message to continue, stopping their current activity and attention for the task they are conducting and activating cognitive resources to process the protective message.

A protective message was expected to reinforce the effect of source information on critical evaluation, which would result in participants evaluating more critically when exposed to an unreliable source. No evidence for this was found. It seems that source information has a greater influence on the critical evaluation of people than a protective message. Perhaps participants were already familiar with the fact that news on Facebook might be unreliable (see also the relatively low score on trust in Facebook news), which made them less involved in processing this message and instead made other cues that give more specific (source) information about whether an actual news message is reliable more important. This might imply that acknowledgement of the source is a more promising intervention when stimulating critical processing is the aim, although more research is needed to verify this.

The second aim of this study was to investigate whether a protective message had an effect on the perception of accuracy in (news) messages on Facebook. Here, the results also show that a protective message has no significant effect. What is striking is that an additional analysis shows that source information can have an undesirable effect on the perception of accuracy of (news) messages that are fake. When disinformation was shown with a reliable source, the news was

considered more accurate than when the disinformation is shown with an unreliable source. The source, thereby, functions as a heuristic for accurate news, while in actuality disinformation is distributed. This study showed that an unreliable source, compared to a reliable source, can cause users to evaluate (news) messages on Facebook more critically. An explanation for this is that people process the reliable source more easily and then trust it more quickly, especially in the social media context that mostly involves quick and casual reading. This results in participants believing a news message from a reliable source more quickly [42,43].

Limitations

This research has several strengths. First, this research makes an important contribution to the social debate about how to deal with online disinformation, by providing scientific evidence on a large and important target group. Second, the stimulus material—a video—attempts to simulate a natural situation on Facebook. The video showed a Facebook timeline with 3 short news messages, comparable to actual news consumption among a large part of the population, increasing the ecological validity of this study. Nonetheless, this research also has a number of limitations. A first limitation of this research is that participants were not asked about the perception or reliability of the chosen news sources. It is unknown whether the participants perceived NOS.nl as a reliable news source, although previous reports have shown this [40], and during the COVID-19 crisis, millions of Dutch people used information from NOS as a source on new developments. A second limitation is that Dailybuzzlive.nl is not characterized by all participants as a news source. Participants may have processed the manipulation but did not interpret Dailybuzzlive.nl as a news source, although this also occurs in real life, where people read items on websites but do not directly consider them as news sources. Third, we only tested the experiment in the Netherlands, so it is difficult to generalize to other countries, particularly in countries where media literacy or eHealth literacy is not high. Considering the importance of the study, this study should be replicated in other countries to see if the outcomes remain the same. Fourth, although our sample size was sufficiently large to detect main effects of news source and warning message (sensitive to detect effect sizes larger than $\eta^2 = .025$ with .80 power), it may have been limited in its capability to detect interaction effects; interaction effect sizes are often smaller than main effect sizes and therefore are harder to detect [46].

Conclusions

This study contributes to the social and scientific debate about how to treat online disinformation. The research shows that a

simple intervention involving the acknowledgement of the source of a (news) message can have a positive effect on the critical evaluation by Facebook users. Furthermore, a protective message does not seem to contribute to the battle against online disinformation. A protective message had no effect on critical evaluation by users and no effect on the perception of accuracy. The results of this research serve as advice for Facebook to show the source more prominently so that users can quickly and easily see where a (news) message is coming from. Source information also has an effect on the perception of accuracy. Strikingly, participants regarded disinformation as more accurate when it came from a reliable news source. This means that reliable news sources, such as the NOS, have a responsibility not to disseminate disinformation, since people are more likely to believe this.

A recommendation for further research is to conduct follow-up studies on source information on Facebook or other social media platforms distributing news in other European countries [47]. Confidence in news is much lower in other European countries than the Netherlands [48]. For example, 59% of people in the Netherlands trust the news, whereas in Italy and France the number is much lower (42% and 35%, respectively). In addition, Facebook is used more as a news source the latter two countries. In the Netherlands, 29% of people use Facebook as a news source, while in Italy and France the numbers are higher (51% and 41%, respectively) [48]. Another recommendation for follow-up research is to conduct the same research with different sources. In this study, we chose the most reliable source in the Netherlands, the NOS [40], and an unreliable source, Dailybuzzlive.nl [40], but it would be very interesting to also test different sources to increase external validity.

At the end of 2016, there was much commotion and media attention around the American presidential elections due to the relatively new phenomenon of online disinformation. Recently, it emerged that millions of Facebook users had seen extreme-right disinformation on the platform in the run-up to the European parliamentary elections [49]. And currently, the European Commission needed to respond to several alternative theories explaining the cause of the COVID-19 crisis because an increasing number of people believe fiction instead of facts [1]. As far as we know, this study was the first to show that source information has an effect on critical evaluation by Dutch Facebook users in the context of online disinformation. This means that critical evaluation can be encouraged when source information is provided, which might help to reduce the negative effects of online disinformation [50,51].

Acknowledgments

We thank Vrije Universiteit Amsterdam for the funding and the opportunity to conduct the research.

Data Availability

We will make the data available upon reasonable request.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-eHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 263 KB - jmir_v24i3e27945_app1.pdf](#)]

References

1. Fighting disinformation. European Commission. 2020. URL: https://ec.europa.eu/info/live-work-travel-eu/health/coronavirus-response/fighting-disinformation_en [accessed 2020-05-24]
2. Lewandowsky S, Ecker UK, Cook J. Beyond misinformation: understanding and coping with the “post-truth” era. *J Appl Res Mem Cogn* 2017 Dec;6(4):353-369. [doi: [10.1016/j.jarmac.2017.07.008](https://doi.org/10.1016/j.jarmac.2017.07.008)]
3. Anderson J, Rainie L. The future of truth and misinformation online. Washington: Pew Internet and American Life Project; 2017 Oct 19. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2017/10/PI_2017.10.19_Future-of-Truth-and-Misinformation_FINAL.pdf [accessed 2020-05-24]
4. Papagiannis G, Posetti J. Journalism, 'fake news' and disinformation: a handbook for journalism education and training. UNESCO. URL: <http://unesdoc.unesco.org/images/0026/002655/265552E.pdf> [accessed 2021-07-18]
5. Marwick A, Lewis R. Media manipulation and disinformation online. Data and Society Research Institute. 2017. URL: https://datasociety.net/wp-content/uploads/2017/05/DataAndSociety_MediaManipulationAndDisinformationOnline-1.pdf [accessed 2022-02-11]
6. Acosta FJ, Rodríguez CJ, Cejas MR, Ramallo-Fariña Y, Fernandez-Garcimartin H. Suicide coverage in the digital press media: adherence to World Health Organization guidelines and effectiveness of different interventions aimed at media professionals. *Health Commun* 2020 Nov;35(13):1623-1632. [doi: [10.1080/10410236.2019.1654176](https://doi.org/10.1080/10410236.2019.1654176)] [Medline: [31469595](https://pubmed.ncbi.nlm.nih.gov/31469595/)]
7. Fletcher R, Cornia A, Graves L, Nielson RK. Measuring the reach of “fake news” and online disinformation in Europe. 2018. URL: <https://tinyurl.com/468fruk6> [accessed 2022-02-11]
8. Salmon CT, Poorisat T. The rise and development of public health communication. *Health Commun* 2020 Nov;35(13):1666-1677. [doi: [10.1080/10410236.2019.1654180](https://doi.org/10.1080/10410236.2019.1654180)] [Medline: [31429307](https://pubmed.ncbi.nlm.nih.gov/31429307/)]
9. Martens B, Aguiar L, Gómez-Herrera E, Muller F. The digital transformation of news media and the rise of disinformation and fake news. 2018. URL: <https://ec.europa.eu/jrc/sites/default/files/jrc111529.pdf> [accessed 2022-02-11]
10. Lee M, Ju Y, You M. The effects of social determinants on public health emergency preparedness mediated by health communication: the 2015 MERS outbreak in South Korea. *Health Commun* 2020 Oct;35(11):1396-1406. [doi: [10.1080/10410236.2019.1636342](https://doi.org/10.1080/10410236.2019.1636342)] [Medline: [31262197](https://pubmed.ncbi.nlm.nih.gov/31262197/)]
11. Allcott H, Gentzkow M. Social media and fake news in the 2016 election. *J Econ Perspect* 2017 May 01;31(2):211-236. [doi: [10.1257/jep.31.2.211](https://doi.org/10.1257/jep.31.2.211)]
12. Guess A, Nyhan B, Reifler J. Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign. 2018. URL: <https://www.american.edu/spa/ccps/politics-of-truth/upload/nyhan-american.pdf> [accessed 2022-02-11]
13. Levine TR. Truth-default theory (TDT): a theory of human deception and deception detection. *J Lang Soc Psychol* 2014 May 23;33(4):378-392. [doi: [10.1177/0261927x14535916](https://doi.org/10.1177/0261927x14535916)]
14. Devaux A, Theben A, Deshpande A, Folkvord F, Meranto A, Porcu F, et al. Study on media literacy and online empowerment issues raised by algorithm-driven media services. European Commission. 2019. URL: <https://op.europa.eu/o/oppportal-service/download-handler?identifier=a9101f97-f940-11e9-8c1f-01aa75ed71a1&format=pdf&language=en&productionSystem=cellar&part=> [accessed 2022-02-11]
15. Chou WS, Hunt YM, Beckjord EB, Moser RP, Hesse BW. Social media use in the United States: implications for health communication. *J Med Internet Res* 2009;11(4):e48 [FREE Full text] [doi: [10.2196/jmir.1249](https://doi.org/10.2196/jmir.1249)] [Medline: [19945947](https://pubmed.ncbi.nlm.nih.gov/19945947/)]
16. Asmolov G. The disconnective power of disinformation campaigns. *J Int Affairs* 2018;71(1.5):69-76 [FREE Full text] [doi: [10.4135/9781483359922.n138](https://doi.org/10.4135/9781483359922.n138)]
17. Shearer E. More than eight-in-ten Americans get news from digital devices. Washington: Pew Internet and American Life Project; 2021 Jan 12. URL: <https://pewrsr.ch/2MZqns7> [accessed 2022-02-11]
18. Walker S, Mercea D, Bastos M. The disinformation landscape and the lockdown of social platforms. *Inform Commun Soc* 2019 Aug 29;22(11):1531-1543. [doi: [10.1080/1369118x.2019.1648536](https://doi.org/10.1080/1369118x.2019.1648536)]
19. Zimmermann F, Kohring M. Mistrust, disinforming news, and vote choice: a panel survey on the origins and consequences of believing disinformation in the 2017 German parliamentary election. *Polit Commun* 2020 Jan 01;37(2):215-237. [doi: [10.1080/10584609.2019.1686095](https://doi.org/10.1080/10584609.2019.1686095)]
20. Burkhardt J. Combating fake news in the digital age. *Library Technol Rep* 2017;53(8):5-9.
21. Zelizer B. Why journalism is about more than digital technology. *Digital Journ* 2019 May 08;7(3):343-350. [doi: [10.1080/21670811.2019.1571932](https://doi.org/10.1080/21670811.2019.1571932)]

22. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* 1974 Sep 27;185(4157):1124-1131. [doi: [10.1126/science.185.4157.1124](https://doi.org/10.1126/science.185.4157.1124)] [Medline: [17835457](https://pubmed.ncbi.nlm.nih.gov/17835457/)]
23. Pennycook G, Epstein Z, Mosleh M, Arechar A, Eckles D, Rand D. Understanding and reducing the spread of misinformation online. *PsyArXiv* :1 Preprint posted online on November 13, 2019. [FREE Full text] [doi: [10.31234/osf.io/3n9u8](https://doi.org/10.31234/osf.io/3n9u8)]
24. Vosoughi S, Roy D, Aral S. The spread of true and false news online. *Science* 2018 Dec 09;359(6380):1146-1151. [doi: [10.1126/science.aap9559](https://doi.org/10.1126/science.aap9559)] [Medline: [29590045](https://pubmed.ncbi.nlm.nih.gov/29590045/)]
25. Bode L, Vraga EK. In related news, that was wrong: the correction of misinformation through related stories functionality in social media. *J Commun* 2015 Jun 23;65(4):619-638. [doi: [10.1111/jcom.12166](https://doi.org/10.1111/jcom.12166)]
26. Bode L, Vraga EK. See something, say something: correction of global health misinformation on social media. *Health Commun* 2018 Sep;33(9):1131-1140. [doi: [10.1080/10410236.2017.1331312](https://doi.org/10.1080/10410236.2017.1331312)] [Medline: [28622038](https://pubmed.ncbi.nlm.nih.gov/28622038/)]
27. Schwarz W, Newman E, Leach N. Making the truth stick & the myths fade: Lessons from cognitive psychology. *Behav Sci Pol* 2016;2(1):85-95. [doi: [10.1353/bsp.2016.0009](https://doi.org/10.1353/bsp.2016.0009)]
28. Burgess J, Hurcombe E. Digital journalism as symptom, response, and agent of change in the platformed media environment. *Digital Journ* 2019 May 08;7(3):359-367. [doi: [10.1080/21670811.2018.1556313](https://doi.org/10.1080/21670811.2018.1556313)]
29. Bringula RP, Catacutan-Bangit AE, Garcia MB, Gonzales JPS, Valderama AMC. "Who is gullible to political disinformation?": predicting susceptibility of university students to fake news. *J Inform Technol Pol* 2021 Jul 07:1-15. [doi: [10.1080/19331681.2021.1945988](https://doi.org/10.1080/19331681.2021.1945988)]
30. Buchanan T. Why do people spread false information online? The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation. *PLoS One* 2020;15(10):e0239666 [FREE Full text] [doi: [10.1371/journal.pone.0239666](https://doi.org/10.1371/journal.pone.0239666)] [Medline: [33027262](https://pubmed.ncbi.nlm.nih.gov/33027262/)]
31. Hameleers M, Powell TE, Van Der Meer TG, Bos L. A picture paints a thousand lies? The effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. *Polit Commun* 2020 Feb 05;37(2):281-301. [doi: [10.1080/10584609.2019.1674979](https://doi.org/10.1080/10584609.2019.1674979)]
32. Chan MS, Jones CR, Hall Jamieson K, Albarracín D. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol Sci* 2017 Nov;28(11):1531-1546 [FREE Full text] [doi: [10.1177/0956797617714579](https://doi.org/10.1177/0956797617714579)] [Medline: [28895452](https://pubmed.ncbi.nlm.nih.gov/28895452/)]
33. Cho J, Park D, Lee HE. Cognitive factors of using health apps: systematic analysis of relationships among health consciousness, health information orientation, eHealth literacy, and health app use efficacy. *J Med Internet Res* 2014;16(5):e125 [FREE Full text] [doi: [10.2196/jmir.3283](https://doi.org/10.2196/jmir.3283)] [Medline: [24824062](https://pubmed.ncbi.nlm.nih.gov/24824062/)]
34. Clayton K, Blair S, Busam JA, Forstner S, Glance J, Green G, et al. Real solutions for fake news? Measuring the effectiveness of general warnings and fact-check tags in reducing belief in false stories on social media. *Polit Behav* 2019 Feb 11;42(4):1073-1095. [doi: [10.1007/s11109-019-09533-0](https://doi.org/10.1007/s11109-019-09533-0)]
35. Wilson SL, Wysong C. Social media and vaccine hesitancy. *BMJ Glob Health* 2020 Oct;5(10):e004206. [doi: [10.1136/bmjgh-2020-004206](https://doi.org/10.1136/bmjgh-2020-004206)] [Medline: [33097547](https://pubmed.ncbi.nlm.nih.gov/33097547/)]
36. Lewandowsky S. Climate change disinformation and how to combat it. *Annu Rev Public Health* 2021 Apr 01;42:1-21. [doi: [10.1146/annurev-publhealth-090419-102409](https://doi.org/10.1146/annurev-publhealth-090419-102409)] [Medline: [33355475](https://pubmed.ncbi.nlm.nih.gov/33355475/)]
37. Dahl S, Eagle L. Empowering or misleading? Online health information provision challenges. *Mark Intell Plan* 2016 Oct 03;34(7):1000-1020. [doi: [10.1108/mip-07-2015-0127](https://doi.org/10.1108/mip-07-2015-0127)]
38. Ekker H. Impact Nederlandse luchtvaart op klimaat neemt flink toe. 2019. URL: <https://nos.nl/artikel/2278632-impact-nederlandse-luchtvaart-op-klimaat-neemt-flink-toe.html> [accessed 2019-04-20]
39. Pennycook G, Rand DG. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc Natl Acad Sci U S A* 2019 Feb 12;116(7):2521-2526 [FREE Full text] [doi: [10.1073/pnas.1806781116](https://doi.org/10.1073/pnas.1806781116)] [Medline: [30692252](https://pubmed.ncbi.nlm.nih.gov/30692252/)]
40. Mediamonitor 2018: Mediabedrijven en mediamarkten. CvdM. 2018. URL: <https://www.cvdM.nl/wp-content/uploads/2018/06/Mediamonitor-2018-Reuters-Institute-Digital-News-Report-Nederland-2018.pdf> [accessed 2019-04-25]
41. Metzger M, Flanagin A, Medders R. Social and heuristic approaches to credibility evaluation online. *J Commun* 2010;60(3):413-439. [doi: [10.1111/j.1460-2466.2010.01488.x](https://doi.org/10.1111/j.1460-2466.2010.01488.x)]
42. Eisend M. Is it still worth to be credible? A meta-analysis of temporal patterns of source credibility effects in marketing. *Adv Consumer Res* 2004;352-357 [FREE Full text]
43. Pornpitakpan C. The persuasiveness of source credibility: A critical review of five decades' evidence. *J Appl Social Psychol* 2004 Feb;34(2):243-281. [doi: [10.1111/j.1559-1816.2004.tb02547.x](https://doi.org/10.1111/j.1559-1816.2004.tb02547.x)]
44. Sherry JL. Flow and media enjoyment. *Commun Theory* 2004 Nov;14(4):328-347. [doi: [10.1111/j.1468-2885.2004.tb00318.x](https://doi.org/10.1111/j.1468-2885.2004.tb00318.x)]
45. Petty R, Cacioppo J. The elaboration likelihood model of persuasion. In: Petty R, Cacioppo J, editors. *Communication and Persuasion*. New York: Springer; 1986:1-24.
46. Perugini M, Gallucci M, Costantini G. A practical primer to power analysis for simple experimental designs. *Int Rev Soc Psychol* 2018;31(1):1. [doi: [10.5334/irsp.181](https://doi.org/10.5334/irsp.181)]
47. Anastasiadou D, Folkvord F, Lupiañez-Villanueva F. A systematic review of mHealth interventions for the support of eating disorders. *Eur Eat Disord Rev* 2018 Sep;26(5):394-416 [FREE Full text] [doi: [10.1002/erv.2609](https://doi.org/10.1002/erv.2609)] [Medline: [29927004](https://pubmed.ncbi.nlm.nih.gov/29927004/)]

48. Reuters Digital News Report 2018. URL: <https://reutersinstitute.politics.ox.ac.uk/sites/default/files/digital-news-report-2018.pdf> [accessed 2019-06-02]
49. Bayer J, Bitiukova N, Bard P, Szakács J. Disinformation and propaganda: impact on the functioning of the rule of law in the EU and its Member States. European Parliament. 2019. URL: [https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU\(2019\)608864_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2019/608864/IPOL_STU(2019)608864_EN.pdf) [accessed 2022-02-18]
50. Potter W. Theory of Media Literacy: A Cognitive Approach. Thousand Oaks: Sage Publications; 2004.
51. Potter W. Media Literacy. Thousand Oaks: Sage Publications; 2018.

Abbreviations

ANCOVA: analysis of covariance

MANCOVA: multivariate analysis of covariance

NOS: Nederlands Omroep Stichting

PCA: principal component analysis

UNESCO: United Nations Educational, Scientific and Cultural Organization

Edited by R Kukařka; submitted 14.02.21; peer-reviewed by M Lotto, M Bardus; comments to author 16.06.21; revised version received 23.07.21; accepted 03.12.21; published 31.03.22.

Please cite as:

Folkvord F, Snelting F, Anschutz D, Hartmann T, Theben A, Gunderson L, Vermeulen I, Lupiáñez-Villanueva F

Effect of Source Type and Protective Message on the Critical Evaluation of News Messages on Facebook: Randomized Controlled Trial in the Netherlands

J Med Internet Res 2022;24(3):e27945

URL: <https://www.jmir.org/2022/3/e27945>

doi: [10.2196/27945](https://doi.org/10.2196/27945)

PMID: [35357314](https://pubmed.ncbi.nlm.nih.gov/35357314/)

©Frans Folkvord, Freek Snelting, Doeschka Anschutz, Tilo Hartmann, Alexandra Theben, Laura Gunderson, Ivar Vermeulen, Francisco Lupiáñez-Villanueva. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 31.03.2022. 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.

Original Paper

Cost-effectiveness of Web-Based and Home-Based Postnatal Psychoeducational Interventions for First-time Mothers: Economic Evaluation Alongside Randomized Controlled Trial

Qishi Zheng^{1,2*}, MPhil; Luming Shi^{1,2,3*}, MBBS, MSc; Lixia Zhu^{4,5*}, PhD; Nana Jiao⁶, MSc; Yap Seng Chong^{7,8}, MBBS, MMED; Sally Wai-Chi Chan⁹, PhD; Yiong Huak Chan¹⁰, PhD; Nan Luo^{5,11}, PhD; Wenru Wang^{4,5}, PhD; Honggu He^{4,5}, PhD

¹Epidemiology, Singapore Clinical Research Institute, Singapore, Singapore

²Cochrane Singapore, Singapore, Singapore

³Duke-NUS Medical School, Singapore, Singapore

⁴Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

⁵National University Health System, Singapore, Singapore

⁶Edson College of Nursing and Health Innovation, Arizona State University, Arizona, AZ, United States

⁷Division of Maternal Fetal Medicine, Department of Obstetrics & Gynaecology, National University Hospital, Singapore, Singapore

⁸Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

⁹Tung Wah College, Hong Kong, China

¹⁰Biostatistics Unit, National University of Singapore, Singapore, Singapore

¹¹Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

*these authors contributed equally

Corresponding Author:

Honggu He, PhD

Alice Lee Centre for Nursing Studies

Yong Loo Lin School of Medicine

National University of Singapore

Level 2, Clinical Research Centre, Block MD11

10 Medical Drive

Singapore, 117597

Singapore

Phone: 65 65167448

Fax: 65 67767135

Email: nurhhg@nus.edu.sg

Abstract

Background: The cost-effectiveness of interventions has attracted increasing interest among researchers. Although web-based and home-based psychoeducational interventions have been developed to improve first-time mothers' postnatal health outcomes, very limited studies have reported their cost-effectiveness.

Objective: The aim of this study was to evaluate the cost-effectiveness of web-based and home-based postnatal psychoeducational interventions for first-time mothers during the early postpartum period.

Methods: A randomized controlled 3-group pretest and posttest design was adopted, and cost-effectiveness analysis from the health care's perspective was conducted. A total of 204 primiparas were recruited from a public tertiary hospital in Singapore from October 2016 to August 2017 who were randomly allocated to the web-based intervention (n=68), home-based intervention (n=68), or control (n=68) groups. Outcomes of maternal parental self-efficacy, social support, postnatal depression, anxiety, and health care resource utilization were measured using valid and reliable instruments at baseline and at 1 month, 3 months, and 6 months after childbirth. The generalized linear regression models on effectiveness and cost were used to assess the incremental cost-effectiveness ratios of the web-based and home-based intervention programs compared to routine care. Projections of cumulative cost over 5 years incurred by the 3 programs at various coverage levels (ie, 10%, 50%, and 100%) were also estimated.

Results: The web-based intervention program dominated the other 2 programs (home-based program and routine care) with the least cost (adjusted costs of SGD 376.50, SGD 457.60, and SGD 417.90 for web-based, home-based, and control group, respectively; SGD 1=USD 0.75) and the best improvements in self-efficacy, social support, and psychological well-being. When considering the implementation of study programs over the next 5 years by multiplying the average cost per first-time mother by the estimated average number of first-time mothers in Singapore during the 5-year projection period, the web-based program was the least costly program at all 3 coverage levels. Based on the 100% coverage, the reduced total cost reached nearly SGD 7.1 million and SGD 11.3 million when compared to control and home-based programs at the end of the fifth year, respectively.

Conclusions: The web-based approach was promisingly cost-effective to deliver the postnatal psychoeducational intervention to first-time mothers and could be adopted by hospitals as postnatal care support.

Trial Registration: ISRCTN registry ISRCTN45202278; <https://www.isrctn.com/ISRCTN45202278>

(*J Med Internet Res* 2022;24(3):e25821) doi:[10.2196/25821](https://doi.org/10.2196/25821)

KEYWORDS

anxiety; cost-effectiveness; depression; first-time mother; home-based; postnatal; psychoeducational; self-efficacy; social support; web-based

Introduction

The early postpartum period is a stressful transition period for first-time mothers owing to physical and emotional challenges related to childbirth and challenges in adapting to their new social roles, for example, parenting newborns and establishing mother-infant relationship [1,2]. Based on the maternal role attainment theory, a mother's emotional health during the postpartum period is mainly influenced by the mother's biopsychosocial well-being, family, and her surrounding environment [3]. Many first-time mothers feel inadequately prepared for motherhood and are situated in environments with insufficient support [4-6].

Psychoeducational interventions have been developed to improve new mothers' knowledge in self-care and newborn care during the postnatal period and to build on their strengths and resources to promote emotional coping and parenting skill development [5,7]. In Asian countries, including Singapore, the follow-up home visits undertaken by midwives are not commonly practiced [8]. Home-based psychoeducation intervention via home visits is not easily accessible owing to the shortage of midwives and nurses, and its cost-effectiveness was undetermined [5,6,8]. Conversely, a similar intervention delivered online, for example, a web-based platform, could address the issues of inaccessibility and high cost [9,10]. About two-thirds of the world population (62%) had access to the internet in 2020 compared to 5.9% in 2000 [11]. In Singapore, a newly initiated Home Access Program allows lower-income households to afford fiber broadband connectivity and a tablet at a subsidized rate to increase the population's internet accessibility [12,13], thus allowing more people to access web-based resources. Moreover, the accessibility to web-based interventions may significantly reduce people's reliance on health care professionals, thus lowering the burden on health care, that is, administrative cost, time, and resource savings [11]. Therefore, if the web-based intervention is effective, it has the potential to be highly cost-effective as it can be delivered at scale across large populations, with relatively low additional costs per additional user [14] unlike face-to-face education or in-person visits where labor costs account for a substantial proportion of the total cost [15].

Although the effectiveness of web-based and mobile app-based interventions [16-18] have been reported, limited information exists on their cost-effectiveness. Previous studies have revealed that compared to routine care or face-to-face intervention, web-based or internet-based intervention could be a cost-effective way for alcohol prevention in adolescents [11], increasing physical activity for healthy adult Latina women [9], dietetic services for weight management [19], and self-management for people with type 2 diabetes [20]. However, there is no study reporting the cost-effectiveness of web-based and home-based psychoeducational interventions for first-time mothers. Such knowledge may help to keep costs to a minimum and make the intervention affordable for all mothers regardless of their socioeconomic status, while achieving positive maternal outcomes.

We conducted a 3-group randomized controlled trial that aimed to examine the effectiveness and cost-effectiveness of web-based and home-based postnatal psychoeducational programs in first-time mothers. The details of the interventions have been reported in the study protocol [12], and the outcomes of their effectiveness have been reported in a separate paper [16]. This paper aims to evaluate the cost-effectiveness of web-based and home-based postnatal psychoeducational interventions on outcomes of self-efficacy in newborn care (primary outcome), social support, postnatal depression, and anxiety (secondary outcomes) in first-time mothers during the early postpartum period.

Methods

Ethics Approval

This study was approved by the National Health Group Domain Specific Review Board on 25 January 2016 (NHG DSRB Ref: 2015/01189).

Study Design, Setting, and Sample

A randomized controlled trial was carried out in a tertiary public hospital in Singapore. First-time mothers recruited from the postnatal wards were randomly allocated to one of the 3 intervention groups: web-based group (receiving web-based psychoeducational intervention plus routine care), home-based

group (receiving home-based psychoeducational intervention plus routine care), or control group (receiving routine care). Mothers were excluded if they had identified physical or mental disorders before and during pregnancy, had complicated assisted delivery with fourth-degree perineal tear, gave birth to a stillborn child, or had a child with a congenital anomaly or medical complications [8]. The sample size of 204 was calculated based on the primary outcome of maternal parental self-efficacy, with 68 allocated in each group [12,16]. The recruitment period was from October 2016 to February 2017, and the last 6 months follow-up data were collected in August 2017. The study design and procedure have been previously reported in papers detailing the study protocol [12] and clinical effectiveness [16]. This study conducted a cost-effectiveness analysis from the health care provider's perspectives alongside the randomized controlled trial.

Interventions

Control Group

Participants allocated to the control group only received routine care provided by the hospital. It involved postnatal support from nurses and midwives in the hospital and a postdelivery follow-up (around 1-6 weeks) outpatient appointment with the doctor. The support focused on providing didactic information on basic baby care tasks while mothers were hospitalized for childbirth as well as the inspection of episiotomy or cesarean section wound and breastfeeding advice during the follow-up hospital visit.

Home-Based Intervention Group

Participants allocated to the home-based group received a 1-hour face-to-face postnatal psychoeducational intervention via a home visit by a registered nurse in addition to routine care. A self-developed educational booklet [12,16] developed based on the evidence of literature by Shorey et al (unpublished data, 2012) and Wong et al [21] was used to facilitate the face-to-face intervention. The booklet was provided to the participants before hospital discharge. Three weekly telephone calls were included in the intervention to address mothers' queries. The postnatal psychoeducational intervention contents were developed based on Bandura's self-efficacy theory [22], social support model, findings from preliminary studies [4,5], and previous literature [23-26].

Web-Based Intervention Group

Briefly, participants in the web-based group received a postnatal psychoeducational intervention via a newly developed website (Figure 1) with 1-month online access after childbirth in addition to routine care. The main contents on the website were identical to those in the booklet, with section-to-section audio files (Figure 2) provided for the mothers' convenience. Videos demonstrating breastfeeding techniques, breast engorgement management, infant bathing, and Kegel exercises based on current practices in local hospitals were also developed and embedded on the website. A peer discussion forum and a confidential corner for the participants to communicate with other mothers or research team members were also available. Three weekly telephone calls, each lasting about 3 minutes, served as reminders for participants to access the website; participants did not receive additional information.

Figure 1. Homepage of the website for mothers in the web-based intervention group.

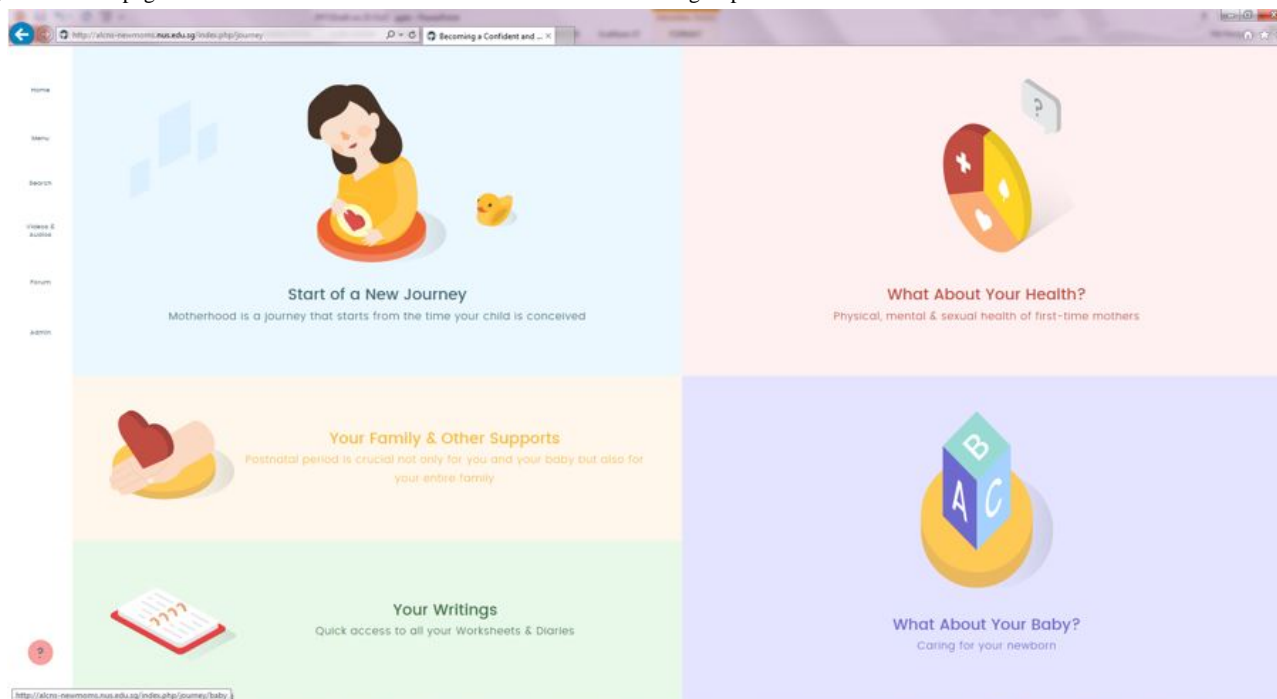


Figure 2. Sample page of voiceover of the contents on the website.

Voiceover

Click on the audio icon to listen to the voiceovers. You can also right-click on the titles to download the files.

All Four Chapters



Full Booklet (01:53:46, 145.1 MB)

Start of a New Journey



Full Chapter



Stories of Other Mothers



A Mother with Breast Engorgement & Breastfeeding Difficulties Learned to be Confident & Competent in Breastfeeding



A Mother Experienced Baby Blues but Overcame with The Help from Health Professionals & Family



A Mother Experiencing Episiotomy Wound Problems with Good Recovery After Care



A Smooth & Successful Childbirth with Happy Early Motherhood Experience with Full Family Support



Your Childbirth Story

Health Outcomes and Instruments

The health outcomes used to measure the effectiveness of the interventions included maternal parental self-efficacy, social support, postnatal depression, and anxiety. The instruments used included the 17-item Perceived Maternal Parental Self-efficacy Scale (total score range 17-68, higher scores meaning better self-efficacy) [27,28], the 22-item modified Perinatal Infant Care Social Support Scale (total score range 22-88, higher scores indicating better social support) [28,29], the 10-item Edinburgh Postnatal Depression Scale (total score range 0-30, lower scores meaning less depressive symptoms) [30], and the Anxiety subscale of the Hospital Anxiety and Depression Scale (total score range 0-21, lower scores meaning less anxious) [31]. The higher scores of self-efficacy and social support and the lower scores of depression and anxiety indicated better outcomes. Detailed descriptions of the instruments and their validity and reliability can be found in papers previously published for this project [12,16].

Costs and Health Care Resource Utilization

The costs referred to in this study were viewed from a health care provider's perspective, which included costs associated with the study interventions and costs related to the health service utilization to treat postnatal medical conditions in mothers and babies. Broadly, costs associated with the study interventions included those incurred during the development and implementation of interventions, ongoing delivery, and maintenance of the interventions. Costs related to the development and implementation of the web-based intervention include the purchasing of goods and services for hardware, software, educational materials, webpage design, maintenance, and work delivered by third-party service providers and were recorded from actual invoices. Costs of delivering the home-based intervention program were captured based on home visit human resources, material costs, and transportation fees.

Costs associated with health service utilization due to postnatal medical conditions were collected using the Questionnaire on Healthcare Service Utilization during 3 periods: within 1 month after childbirth, 1-3 months after childbirth, and 3-6 months

after childbirth. All medical conditions that required health care services were captured and classified into 2 categories: infant-related conditions and maternal-related conditions. Service provider information was also recorded, which included government polyclinics, private clinics (including general practitioner clinics), hospital specialist outpatient clinics, hospital emergency departments, hospital inpatient admissions, over-the-counter pharmacies, and others. For every health service received by patients, data on full bills (ie, before any subsidies or insurance payments) were collected to estimate the health care provider's full cost for each service. In Singapore, the public health care delivery system is not-for-profit. The entire bill size is the best estimate for the true cost incurred for delivering care. The entire bill size for certain health care services is not charged at a flat rate; it varies among patients depending on the duration of care delivery at each visit and the seniority of the personnel attending to the patient. For example, the cost of an outpatient visit is higher if the consultation time with a senior consultant is longer compared to a shorter consultation time with a junior consultant, although the unit cost for the medications and procedures is standardized.

Data Analysis

Costs Associated With Study Interventions

Costs related to development and implementation of the web-based intervention were further converted to an annualized average cost per mother by assuming that the website could be used for 5 years [32]. We also considered a depreciation factor at 3% for the calculation of the above costs. Costs related to delivering home-based interventions were averaged for each participating patient in the study group as well. Since all 3 groups received identical routine care, the costs associated with standard care were not included in the cost-effectiveness analysis.

Costs Associated With Postnatal Health Services

For postnatal health services related to mothers' and infants' medical conditions, the costs were summarized into 3 categories: service providers, infant-related conditions, and maternal-related conditions. For each health service, for example, the general practitioner and private clinic, the average cost per visit was estimated by dividing the total bills incurred from patients who utilized this service during the study period by the total number of such visits. All costs were calculated in SGD (2016-2017, SGD 1=USD 0.75).

Base Case Analysis

The base case analysis included all participants with complete follow-up information and original randomized allocation (N=193). The generalized linear regression models on effectiveness and cost at the sixth month were applied to assess the cost-effectiveness of the 3 programs, adjusting for relevant sociodemographic and clinical characteristics. Univariable models were first fit to identify the potential confounding factors associated with effectiveness and cost. For example, the effectiveness model was adjusted for potential confounding factors such as baseline values, age, number of days since baseline, ethnicity, employment status, monthly household income, prenatal courses attendance, and skin-to-skin contact

with baby. The cost model was adjusted for factors that might influence the utilization of health and other services (ie, employment status and income levels). These adjustments aimed to obtain average effectiveness and cost accounting for the influence from abovementioned risk factors instead of the crude results that might not be representative. To account for mothers with zero cost, that is, those who did not consume any health care resources during the study period, a 2-part model approach was adopted. The first-part probit model estimated the probability of zero cost and the second-part generalized linear regression model, with a log link in gamma family, handled nonzero positive cost data. For the base case analysis of cost-effectiveness, time horizon for the health outcomes and costs associated with ongoing delivery and maintenance of the interventions were consistent with the randomized controlled trial, that is, 6 months, with no discounting applied. Incremental cost-effectiveness ratios (ICERs) were estimated to measure the economic value of the study interventions in comparison to the control group. It was calculated by dividing the difference in total costs (incremental cost) by the difference in health outcomes or effects at 6-month postintervention (incremental effect) to provide a ratio of extra cost per unit of incremental health effect. When calculating the ICER, the depression and anxiety scores were inverted, which means that the higher the score, the better after inversion.

Sensitivity Analyses

Probabilistic sensitivity analyses were adopted to address the parameter uncertainty, that is, cost and effectiveness. The distribution of the resulting 10,000 estimates of the ICER on the cost-effectiveness plane depicts a joint uncertainty surrounding costs and outcomes. Estimates in the lower right quadrant of a cost-effectiveness plane suggest that the program is more effective and less costly, while estimates in the upper right quadrant suggest that it is more effective but costlier. The cost-effectiveness acceptability curve was used to display the probability that 1 program is cost-effective at a given willingness-to-pay threshold. Several scenario analyses were conducted to re-estimate the ICER to explore the results' sensitivity to particular assumptions or parameters: (1) to assume web-based and home-based program would be used for only 3 years and (2) to vary the overall program cost from 0% to 300%.

Projection of Nationwide Implementation of Interventions Over 5 Years

The economic impact of national implementation of study interventions over the next 5 years was estimated by multiplying the average cost per the first-time mother by the estimated average number of first-time mothers in Singapore. The number of first-order babies born from 2012 to 2016 in Singapore ranged from 19,292 to 20,755, yielding an average of 20,000 first-order babies each year [33]. In our projection, we assumed that all 20,000 first-time mothers would be eligible for the psychoeducational program. To consider different acceptance levels of interventions, 3 program coverage levels, that is, 10%, 50%, and 100% for the eligible first-time mothers were adopted for economic impact analysis.

Results

Characteristics of the Participants and the Baseline Health Outcomes Among the 3 Groups

In total, 204 mothers were recruited, with 193 included in the economic evaluation (64 for web-based, 63 for home-based, and 66 for control group). The CONSORT diagram and details of the mothers' characteristics have previously been published [16]. Briefly, the mean age of the mothers was around 30 years for all 3 groups, and half of them were of Chinese ethnicity. Over 80% of the first-time mothers obtained a bachelor's degree or above (165/204) and were employed (175/204, 85%) during the study period. Nearly three-fourths of the mothers had monthly household incomes of more than SGD 5000 and the majority (178/204, 87%) were healthy and had no chronic diseases. Overall, around 58% (119/204) of the first-time mothers had normal deliveries. There were no statistically significant differences in all sociodemographic, clinical, and baseline data among mothers in the 3 groups. There were no significant differences in the baseline outcomes of self-efficacy, social support, postnatal depression, and anxiety among the 3 groups [16].

Costs

During the 6-month follow-up period, health services utilization data seeking various services concerning infant-related and maternal-related conditions showed similar occurrence rates among the 3 groups (Table 1). The number and percentage of zero-health service use mothers were as follows: web-based (28/64 44%), home-based (20/63, 32%), and control (30/66, 46%). The average health care cost per participant was SGD 339.90 (SD 493) for the web-based group, SGD 375.40 (SD 699.9) for the home-based group, and SGD 411.20 (SD 601) for the control group. Moreover, the web-based program involved an upfront cost of SGD 17,210 to develop both the web design and educational videos and an annual web maintenance fee of SGD 2900 per year. Similarly, the home-based program required a SGD 233.40 upfront one-time training fee for the payment of an experienced Registered Midwife to train the Registered Nurse who delivered the face-to-face postnatal psychoeducation; home visits cost SGD 77.42 for 1-hour session, which is inclusive of the manpower cost as well as the costs of printed materials and transportation. When accounting for all relevant costs, including 6 months' health care utilization cost, the web-based program cost an average of SGD 390.80/participant, while home-based and control programs cost SGD 437.80/participant and SGD 411.20/participant, respectively.

Table 1. Service use during the follow-up period (N=193).

	Web-based intervention (n=64) ^a		Home-based intervention (n=63) ^b		Control (routine care) (n=66) ^c	
	n	Cost (SGD) ^d , mean (SD)	n	Cost (SGD), mean (SD)	n	Cost (SGD), mean (SD)
Service providers						
General practitioner and private clinic	43	127 (99.28)	44	150 (171.72)	41	260 (400.52)
Hospital outpatient	17	205 (201.90)	21	264 (241.48)	34	208 (193.93)
Hospital accident and emergency	25	199 (197.84)	24	392 (990.39)	16	276 (600.57)
Hospital inpatient	6	771 (452.60)	2	200 (0)	8	968 (744.45)
Polyclinic	2	149 (148.49)	0	0	4	38 (25.63)
Over-the-counter	0	0 (0)	2	18 (2.83)	1	20 (0)
Others	19	179 (185.07)	13	92 (67.15)	6	155 (163.28)
Infant-related conditions						
Spit up (milk)	12	196 (125.84)	6	198 (174.98)	13	110 (108.66)
Excessive eye discharge	5	156 (61.81)	1	55 (0)	2	519 (536.23)
Infantile eczema	3	76 (70.85)	5	96 (78.46)	12	217 (270.02)
Diaper rash	4	74 (36.77)	1	34 (0)	4	180 (195.25)
Others	48	239 (287.20)	54	244 (613.23)	43	298 (492.38)
Total	72	215 (248.68)	67	248 (617.18)	74	255 (409.74)
Maternal-related conditions						
Insufficient milk supply	13	128 (192.12)	8	100 (83.47)	5	91 (30.61)
Breast engorgement	9	139 (114.95)	7	145 (98.33)	7	175 (144.01)
Mastitis	3	94 (70.06)	10	282 (351.09)	2	56 (61.77)
Prolonged lochia	1	30 (0)	0	0 (0)	1	20 (0)
Others	14	219 (231.55)	15	179 (134.66)	21	351 (496.33)
Total	40	157 (186.14)	40	183 (204.62)	36	257 (403.54)

^aUpfront cost: SGD 17,210, follow-up cost: SGD 2900 per year, average total cost per participant (a depreciation factor 3% over 5 years was considered for the average total cost per participant calculation): SGD 390.8, average health care cost: SGD 339.9 (SD 493).

^bUpfront cost: SGD 233.4, follow-up cost: SGD 77.42 per visit, average total cost per participant (a depreciation factor 3% over 5 years was considered for the average total cost per participant calculation): SGD 437.8, average health care cost: SGD 375.4 (SD 699.9).

^cUpfront cost: SGD 0, follow-up cost: SGD 0, average total cost per participant (a depreciation factor 3% over 5 years was considered for the average total cost per participant calculation): SGD 411.2, average health care cost: SGD 411.2 (SD 601).

^dSGD 1=USD 0.75.

Cost-effectiveness Analyses Results

The effectiveness of the web-based and home-based psychoeducational intervention programs has previously been reported [16]. Point estimates of the ICER for the scores of self-efficacy, social support, postnatal depression, and anxiety are reported in Table 2. In Table 2, the adjusted effectiveness of each outcome (self-efficacy, social support, depression, and anxiety) was estimated by adjusting for baseline values, age, number of days since the baseline, ethnicity, employment status, monthly household income, prenatal courses attendance, and skin-to-skin contact with the baby. The base-case analysis showed that the web-based group incurred the least cost and was the most effective among the 3 groups across all outcome measures. It saved around SGD 81 per participant (18% of the average adjusted cost) compared to the home-based program

(adjusted cost: web-based group SGD 376.50 vs home-based group SGD 457.60) while showing better performance in all the outcomes, and SGD 41 per participant compared to the control group (adjusted cost: web-based group SGD 376.50 vs control group SGD 417.90) with better performance in all the outcomes. Hence, it was considered dominant across all the outcomes. When comparing the home-based group to the control group, the ICER ranged from SGD 8.60 per social support score improvement to SGD 90.20 per anxiety score (Anxiety subscale of the Hospital Anxiety and Depression Scale) reduction. When the validity of the programs was set for 3 years, the web-based program remained the dominant strategy to provide postnatal care (Table 2). The same pattern was observed when the overall program cost was varied at 0% and 50%. At 200% and 300% of the original overall program costs, the control program

became the most cost-effective strategy, followed by the web-based and home-based programs.

A probabilistic sensitivity analysis was adopted to account for uncertainty from the trial. Cost-effectiveness planes comparing web-based versus home-based and home-based versus control were also generated (Figures 3 and 4). As shown in Figure 3, the ICER estimates for self-efficacy and social support were more in the right-side quadrants when comparing web-based versus home-based interventions, and the ICER estimates for maternal depression and anxiety were more in the left-side quadrants. Similar patterns were observed for the comparison of home-based versus control groups (Figure 4). The cost-effectiveness acceptability curves (Figure 5) showed the probability of cost-effectiveness for the 3 groups at various willingness-to-pay levels for the scores of self-efficacy, social

support, postnatal depression, and anxiety. The pattern of the cost-effectiveness acceptability curves for the self-efficacy, social support, postnatal depression, and anxiety scores were similar, with the highest curve from the web-based program and the lowest curve from the control program. As the web-based program was the most effective and was associated with the lowest average total cost among the 3 programs, it was most likely to be cost-effective at all willingness-to-pay levels. However, its probability of cost-effectiveness showed a flat trend when the willingness-to-pay level reached around SGD 100 per self-efficacy or per social support (per score change on the scale) and SGD 400 per maternal depression or per maternal anxiety (per score change on the scale). The lowest curve for all 4 outcomes was for the control group, which indicated that the routine care was the least likely to be cost-effective than the other 2 intervention programs.

Table 2. Incremental cost-effectiveness ratios from the base case and different scenarios for various outcome measures (N=193).

Scenario, group	Adjusted cost (SGD) ^a	Self-efficacy (PMPSE) ^b		Social support (PICSS-modified) ^c		Depression (EPDS) ^d		Anxiety (HADS-A) ^e		
		AE ^f	ICER ^g	AE	ICER	AE	ICER ^h	AE	ICER ^h	
Base case										
W ^{i,j}	376.50	59.11	Dominant ^k	81.6	Dominant	4.44	Dominant	3.36	Dominant	
C ^l	417.90	56.28	N/A ^m	75.3	N/A	6.17	N/A	4.81	N/A	
H ⁿ	457.60	57.27	40.1	79.9	8.6	5.45	55.1	4.37	90.2	
Program valid for 3 years only										
W ⁱ	392.90	59.11	Dominant	81.6	Dominant	4.44	Dominant	3.36	Dominant	
C	417.90	56.28	N/A	75.3	N/A	6.17	N/A	4.81	N/A	
H	458.20	57.27	40.7	79.9	8.8	5.45	56.0	4.37	91.6	
0% of the overall program cost										
W ⁱ	325.55	59.11	Dominant	81.6	Dominant	4.44	Dominant	3.36	Dominant	
H	380.22	56.28	Dominated ^o	75.3	Dominated	6.17	Dominated	4.81	Dominated	
C	417.90	57.27	Dominated	79.9	Dominated	5.45	Dominated	4.37	Dominated	
50% of the overall program cost										
W ⁱ	351.01	59.11	Dominant	81.6	Dominant	4.44	Dominant	3.36	Dominant	
C	417.90	56.28	Dominated	75.3	Dominated	6.17	Dominated	4.81	Dominated	
H	419.12	57.27	Dominated	79.9	Dominated	5.45	Dominated	4.37	Dominated	
200% of the overall program cost										
C ⁱ	417.90	59.11	N/A	81.6	N/A	4.44	N/A	3.36	N/A	
W	427.38	56.28	3.3	75.3	1.5	6.17	5.5	4.81	6.5	
H	535.81	57.27	Dominated	79.9	Dominated	5.45	Dominated	4.37	Dominated	
300% of the overall program cost										
C	417.90	59.11	N/A	81.6	N/A	4.44	N/A	3.36	N/A	
W	478.29	56.28	21.3	75.3	9.6	6.17	34.9	4.81	41.6	
H	613.61	57.27	Dominated	79.9	Dominated	5.45	Dominated	4.37	Dominated	

^aSGD 1=USD 0.75.

^bPMPSE: Perceived Maternal Parental Self-Efficacy.

^cPICSS-modified: modified Perinatal Infant Care Social Support.

^dEPDS: Edinburgh Postnatal Depression Scale.

^eHADS-A: Anxiety subscale of the Hospital Anxiety and Depression Scale.

^fAE: adjusted effectiveness.

^gICER: incremental cost-effectiveness ratio.

^hThe scores of depression/anxiety were inverted before calculating the incremental cost-effectiveness ratio so that the higher scores indicate the better after inversion.

ⁱThe group with the least cost in each scenario is italicized.

^jW: web-based intervention group.

^kDominant means the program is the best in both cost and efficacy among the 3 groups.

^lC: control group.

^mN/A: not applicable.

ⁿH: home-based intervention group.

^oDominated means the program is the worst in both cost and efficacy among the 3 groups.

Figure 3. Cost-effectiveness planes of web-based versus home-based groups (home-based as the base for the incremental calculation) for all outcomes (favored the positive side for all outcomes, the scores of depression and anxiety had been inverted when calculating the incremental cost-effectiveness ratio) (N=193).

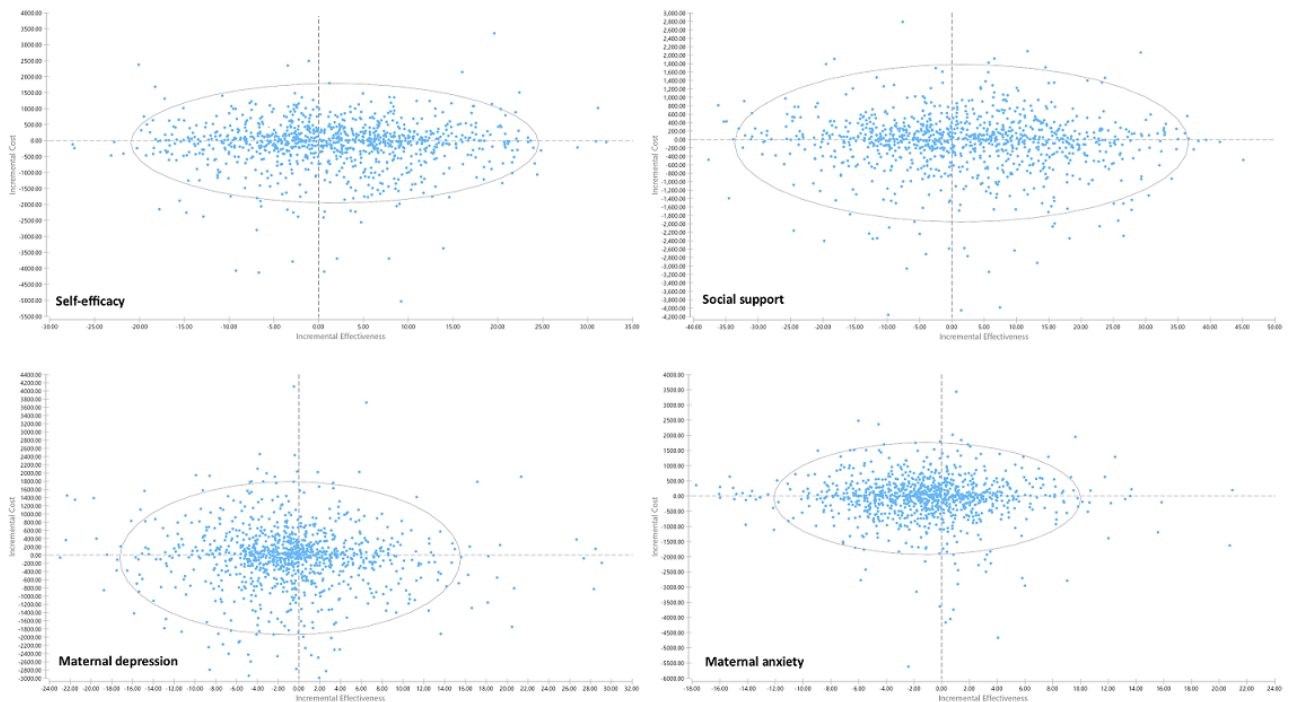


Figure 4. Cost-effectiveness planes of home-based versus control groups (control as the base for the incremental calculation) for all outcomes (favored the positive side for all outcomes, the scores of depression and anxiety had been inverted when calculating the incremental cost-effectiveness ratio) (N=193).

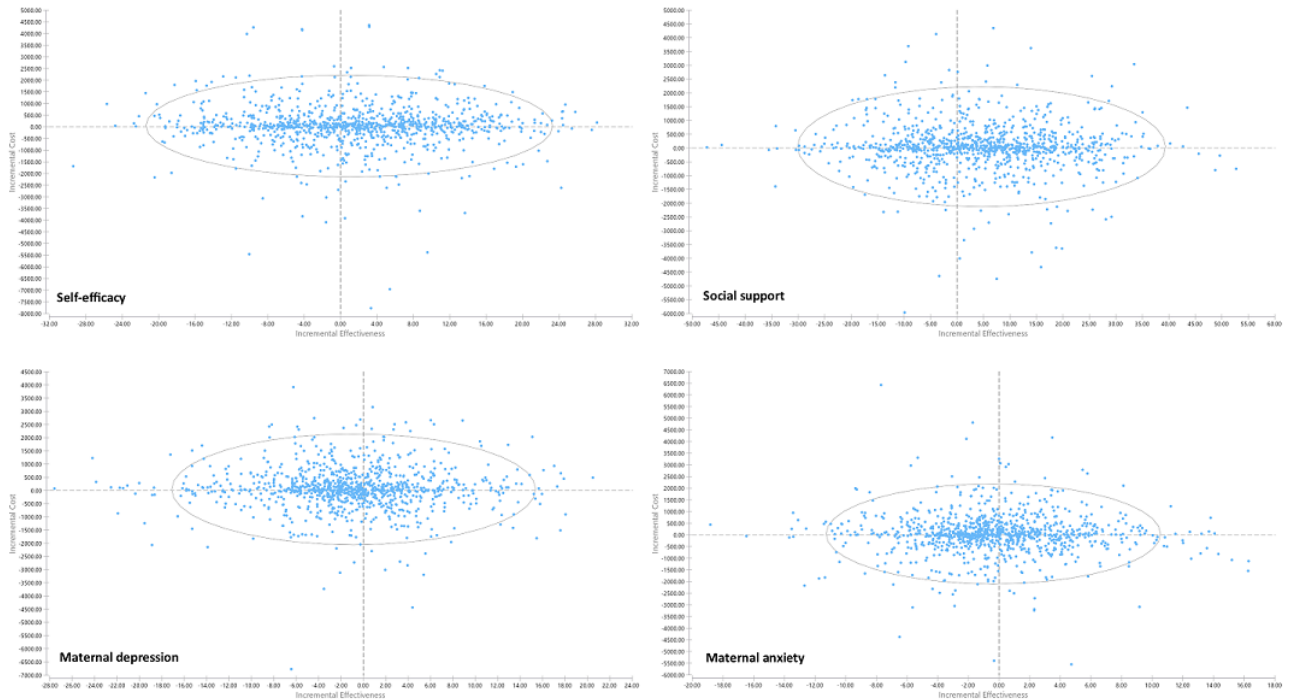
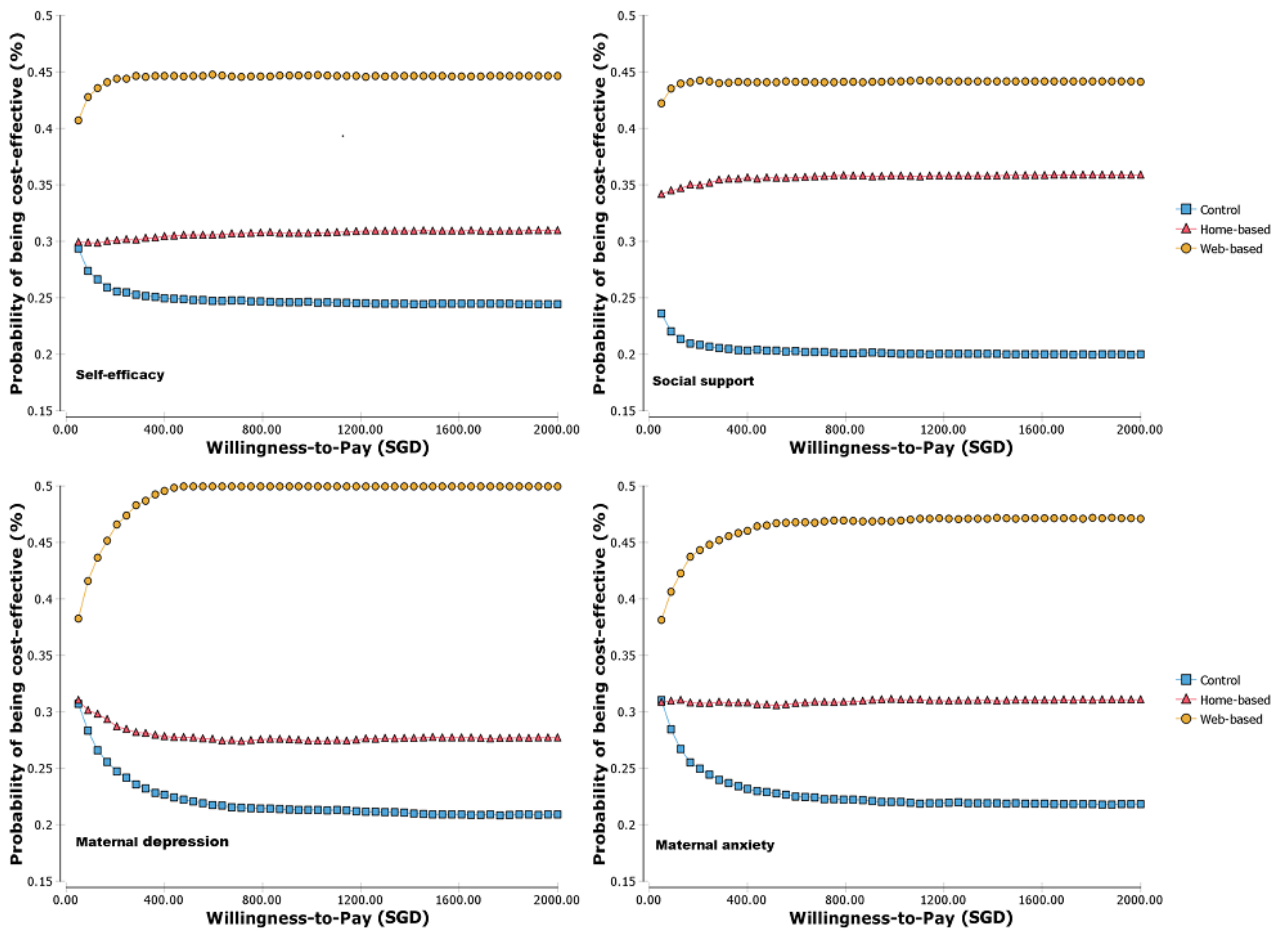


Figure 5. Cost-effectiveness acceptability curves for all outcomes at various willingness-to-pay thresholds (per score improvement) among the 3 groups (N=193).



Projection of Nationwide Implementation of Interventions Over 5 Years

As shown in Table 3, during the 5-year projection period, the web-based intervention was the least costly program at all 3 coverage levels (ie, SGD 3,430,711 at 10% coverage, SGD 17,026,711 at 50%, and SGD 34,021,711 at 100%), and based on the 100% coverage, the reduced total cost reached nearly SGD 7.1 million (41,120,000 minus 34,021,711) and SGD 11.3 million (45,282,233 minus 34,021,711) compared to the control and home-based groups at the end of the fifth year, respectively.

Although there was a relatively high upfront cost for the web-based program, it needed a very low maintenance fee for each year after the first year and provided almost unlimited concurrent access to the web-based contents for all target populations. On the contrary, the home-based program required intensive manpower to deliver face-to-face teaching for each visit, which was sensitive to the program’s coverage and comprised 17% of the total cost. Therefore, even with the lower health care cost over 5 years, the home-based program was the costliest in terms of the total cost incurred based on the base-case analysis results.

Table 3. Projections of cumulative cost over 5 years incurred by 3 programs at different coverage levels for first-time mothers (N=193).^a

Group, cost, coverage	Upfront (SGD)	First year (SGD)	Second year (SGD)	Third year (SGD)	Fourth year (SGD)	Fifth year (SGD)
Web-based intervention group						
Program cost						
10%, 50%, 100%	17,210	20,110	23,011	25,911	28,811	31,711
Health care cost						
10%	0	679,800	1,359,600	2,039,400	2,719,200	3,399,000
50%	0	3,399,000	6,798,000	10,197,000	13,596,000	16,995,000
100%	0	6,798,000	13,596,000	20,394,000	27,192,000	33,990,000
Total cost						
10%	17,210	699,910	1,382,611	2,065,311	2,748,011	3,430,711
50%	17,210	3,419,110	6,821,011	10,222,911	13,624,811	17,026,711
100%	17,210	6,818,110	13,619,011	20,419,911	27,220,811	34,021,711
Home-based intervention group						
Program cost						
10%	233	155,073	309,913	464,753	619,593	774,433
50%	233	774,433	1,548,633	2,322,833	3,097,033	3,871,233
100%	233	1,548,633	3,097,033	4,645,433	6,193,833	7,742,233
Health care cost						
10%	0	750,800	1,501,600	2,252,400	3,003,200	3,754,000
50%	0	3,754,000	7,508,000	11,262,000	15,016,000	18,770,000
100%	0	7,508,000	15,016,000	22,524,000	30,032,000	37,540,000
Total cost						
10%	233	905,873	1,811,513	2,717,153	3,622,793	4,528,433
50%	233	4,528,433	9,056,633	13,584,833	18,113,033	22,641,233
100%	233	9,056,633	18,113,033	27,169,433	36,225,833	45,282,233
Control group						
Program cost						
10%, 50%, 100%	0	0	0	0	0	0
Health care cost						
10%	0	822,400	1,644,800	2,467,200	3,289,600	4,112,000
50%	0	4,112,000	8,224,000	12,336,000	16,448,000	20,560,000
100%	0	8,224,000	16,448,000	24,672,000	32,896,000	41,120,000
Total cost						
10%	0	822,400	1,644,800	2,467,200	3,289,600	4,112,000
50%	0	4,112,000	8,224,000	12,336,000	16,448,000	20,560,000
100%	0	8,224,000	16,448,000	24,672,000	32,896,000	41,120,000

^aSGD 1=USD 0.75.

Discussion

Principal Results

To the best of our knowledge, this is the first study that economically evaluated modes of delivering psychoeducation to first-time mothers. Data on economic outcome measures were

less frequently reported [34]. Findings from this study indicated that the web-based psychoeducation program was the most cost-effective approach compared to the home-based and control programs. In the base-case analysis, the web-based psychoeducation program dominated the other 2 programs across all outcomes. It saved around SGD 81 per participant (18% of the average adjusted cost) compared to the home-based program

and around SGD 41 per participant compared to the routine care alone while showing better performance over them. Our findings are consistent with the results from previous studies demonstrating that web-based or internet-based intervention is cost-effective for various populations when compared to routine care or home-based intervention [10,11,19,20]. In addition, our findings are in line with results from systematic reviews and meta-analyses of the effectiveness and cost-effectiveness of eHealth interventions in somatic diseases, in which most studies indicated that eHealth was effective and cost-effective or at least promising [34]. Another systematic review and meta-analysis reporting the effectiveness and cost-effectiveness of eHealth interventions for depression and anxiety in primary care demonstrated that eHealth was only cost-effective for depression, but there was no evidence for its effectiveness for anxiety [35]. The findings are partly inconsistent with ours, which demonstrated that the web-based program was more cost-effective than the home-based and control programs for self-efficacy, social support, depression, and anxiety. In our sensitivity analysis, we also noticed that the probability of being the best cost-effective was between 45% and 50% for all 4 reported outcomes (Figure 5). Hence, our findings provided evidence that the web-based approach may be a cost-effective way of delivering psychoeducational interventions but need further confirmation with large scale data (eg, real-world study) [36].

To evaluate the budget impact of adopting the programs, we used the average cost data and the estimated number of first-time mothers in Singapore to project the cumulative cost over 5 years [11]. Assuming a 10% to 100% coverage of the web-based psychoeducation program, we found that around SGD 1 million to SGD 11.2 million would be saved compared to the home-based program or around SGD 680,000 to SGD 7.1 million costs would be saved when compared to the control program at the fifth year. The tremendous projected savings for the web-based program make it promising and attractive to be implemented nationwide. In short, over the 5-year projection, the web-based program showed long-term cost-savings compared to the home-based program and the routine care. Although the upfront cost was high for the web-based program, the low maintenance fee made it superior to the other 2 groups.

Strengths and Limitations

This study has several strengths, which include its rigorous study design and the use of market data for personnel costs. The cost-effectiveness analysis was conducted alongside the 3-group repeated measures randomized controlled trial with a 6-month follow-up period. To evaluate the robustness of the results, we conducted both deterministic and probabilistic sensitivity analyses to check the uncertainty of our model. When the overall program cost was increased by 2-3 times (200% and 300% of the overall program costs, respectively), the control group cost the least among the 3 programs. However, the ICER of the web-based program over control program was rather low, that is, ranging from SGD 1.50 to SGD 6.50 per score improvement among all outcomes for 200% of the overall program cost and from SGD 9.60 to SGD 41.60 per score improvement among all outcomes for 300% of the overall program cost. The web-based program could be a preferred approach, given these

relatively low ICERs. The cost-effectiveness acceptability curves were constructed to assess the sampling uncertainty from the trial based on 10,000 times bootstrapped results. The probabilities of the web-based program being cost-effective increased when the willingness-to-pay threshold increased and remained the highest among all 3 groups regardless of the willingness-to-pay thresholds. Despite the strengths, there were some limitations in this study. First, medical utilization data were collected at an individual level for all participants during the follow-up period and captured all possible costs incurred. However, the sample size was estimated based on the clinical effectiveness of the primary outcome of self-efficacy rather than the cost-effectiveness; thus, this might have contributed to uncertainty surrounding the results [37]. Meanwhile, there were 11 participants with missing cost data for the analysis, which might have affected the estimation of the cost for each group. To better address the uncertainty, we conducted a series of deterministic and probabilistic sensitivity analyses and found that the results were quite consistent and robust across the various analyses. In addition, this study used a randomized controlled trial with a 3-group study design, and the effectiveness of the interventions was evaluated up to 6 months after childbirth (about 5 months after the intervention), thus contributing to the robust study design [12,16] and ensuring that the findings from this study were representative and meaningful. Second, this analysis estimated costs from a health care perspective and not the estimated costs from a societal perspective. Therefore, it is recommended that future studies should conduct cost-effectiveness analysis of similar intervention from a social perspective. Third, the measurements of the health services utilization were based on self-reports at each follow-up data collection, which might lead to an underestimation of health resource use in comparison with daily recording in diaries, as people may forget to recall the services used [10]. However, in this study, the recall periods were kept short. The recall period for the costs of health resource used was only 3 months. Moreover, since we used a randomized design in this study, this underestimation was likely to be equally distributed among the 3 groups. Therefore, it is unlikely that the ICERS were affected in this study. Lastly, the calculated costs of the projection of nationwide implementation of interventions over 5 years were based on all first-time mothers, including those with complications. Since we excluded these mothers with severe complications in this study, our estimated costs might be conservative since the actual health care costs for those with complications might be higher. However, this would not affect the group comparison outcomes.

Clinical Implications and Recommendations for Future Studies

Costs, clinical considerations, and participant preferences are often used to determine the choice of a treatment or intervention and its delivery [38]. The findings from this study suggest that a web-based postnatal psychoeducational intervention for first-time mothers was the most cost-effective approach for all outcomes measured. Currently, such postnatal psychoeducational interventions are not routinely provided in Singapore and need to be seriously considered in postnatal care to support mothers, especially first-time mothers. In addition, future studies using

diaries to record the cost of using health resources and estimating costs from both health care and societal perspectives are recommended.

Conclusions

Our findings suggest that compared to home-based psychoeducational intervention and routine care, the web-based approach was promisingly cost-effective to deliver the postnatal psychoeducational intervention to first-time mothers and could be adopted by hospitals as part of postnatal care support.

Acknowledgments

The authors appreciate the great support provided by the nursing managers and staff at postnatal wards of the National University of Singapore. We would like to thank Dr Shefaly Shorey for her contribution to this study. We acknowledge Dr Jennie Wong of Medical and Scientific Communication, Research Support Unit, National University Health System for editing this manuscript.

This study received the National Medical Research Council Health Services Research Competitive Research Grant from the Ministry of Health, Singapore (Project No.: NMRC/HSRG/0051/2015).

Conflicts of Interest

None declared.

References

1. Cowan CP, Cowan PA. *When Partners Become Parents: The Big Life Change for Couples*. Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
2. Murray L, Arteche A, Fearon P, Halligan S, Goodyer I, Cooper P. Maternal postnatal depression and the development of depression in offspring up to 16 years of age. *J Am Acad Child Adolesc Psychiatry* 2011 May;50(5):460-470. [doi: [10.1016/j.jaac.2011.02.001](https://doi.org/10.1016/j.jaac.2011.02.001)] [Medline: [21515195](https://pubmed.ncbi.nlm.nih.gov/21515195/)]
3. Rubin R. Attainment of the maternal role. Part I: Processes. *Nurs Res* 1967;16(3):237-245. [doi: [10.1097/00006199-196701630-00006](https://doi.org/10.1097/00006199-196701630-00006)]
4. Forster DA, McLachlan HL, Rayner J, Yelland J, Gold L, Rayner S. The early postnatal period: exploring women's views, expectations and experiences of care using focus groups in Victoria, Australia. *BMC Pregnancy Childbirth* 2008 Jul 22;8(1):27 [FREE Full text] [doi: [10.1186/1471-2393-8-27](https://doi.org/10.1186/1471-2393-8-27)] [Medline: [18644157](https://pubmed.ncbi.nlm.nih.gov/18644157/)]
5. Ong SF, Chan WS, Shorey S, Chong YS, Klainin-Yobas P, He H. Postnatal experiences and support needs of first-time mothers in Singapore: a descriptive qualitative study. *Midwifery* 2014 Jun;30(6):772-778. [doi: [10.1016/j.midw.2013.09.004](https://doi.org/10.1016/j.midw.2013.09.004)] [Medline: [24161493](https://pubmed.ncbi.nlm.nih.gov/24161493/)]
6. Shorey S, Chan SWC, Chong YS, He H. A randomized controlled trial of the effectiveness of a postnatal psychoeducation programme on self-efficacy, social support and postnatal depression among primiparas. *J Adv Nurs* 2015 Jun;71(6):1260-1273. [doi: [10.1111/jan.12590](https://doi.org/10.1111/jan.12590)] [Medline: [25496615](https://pubmed.ncbi.nlm.nih.gov/25496615/)]
7. Rowe HJ, Fisher JR. Development of a universal psycho-educational intervention to prevent common postpartum mental disorders in primiparous women: a multiple method approach. *BMC Public Health* 2010 Aug 18;10(1):499 [FREE Full text] [doi: [10.1186/1471-2458-10-499](https://doi.org/10.1186/1471-2458-10-499)] [Medline: [20718991](https://pubmed.ncbi.nlm.nih.gov/20718991/)]
8. Shorey S, Chan WS, Chong YS, He H. A randomized controlled trial of the effectiveness of a postnatal psychoeducation programme on outcomes of primiparas: study protocol. *J Adv Nurs* 2015 Jan;71(1):193-203. [doi: [10.1111/jan.12461](https://doi.org/10.1111/jan.12461)] [Medline: [24909836](https://pubmed.ncbi.nlm.nih.gov/24909836/)]
9. Drost RMWA, Paulus ATG, Jander AF, Mercken L, de Vries H, Ruwaard D, et al. A web-based computer-tailored alcohol prevention program for adolescents: Cost-effectiveness and intersectoral costs and benefits. *J Med Internet Res* 2016 Apr 21;18(4):e93 [FREE Full text] [doi: [10.2196/jmir.5223](https://doi.org/10.2196/jmir.5223)] [Medline: [27103154](https://pubmed.ncbi.nlm.nih.gov/27103154/)]
10. Larsen B, Marcus B, Pekmezi D, Hartman S, Gilmer T. A web-based physical activity intervention for Spanish-speaking Latinas: A costs and cost-effectiveness analysis. *J Med Internet Res* 2017 Feb 22;19(2):e43 [FREE Full text] [doi: [10.2196/jmir.6257](https://doi.org/10.2196/jmir.6257)] [Medline: [28228368](https://pubmed.ncbi.nlm.nih.gov/28228368/)]
11. Internet users in the world. Internet World Stats. URL: <http://www.internetworldstats.com/stats.htm> [accessed 2020-09-11]
12. He H, Zhu L, Chan SWC, Chong Y, Jiao N, Chan YH, et al. The Effectiveness and Cost-Effectiveness of Web-Based and Home-Based Postnatal Psychoeducational Interventions for First-Time Mothers: Randomized Controlled Trial Protocol. *JMIR Res Protoc* 2018 Jan 31;7(1):e35 [FREE Full text] [doi: [10.2196/resprot.9042](https://doi.org/10.2196/resprot.9042)] [Medline: [29386175](https://pubmed.ncbi.nlm.nih.gov/29386175/)]
13. Home access programme to benefit 8,000 low income households. Infocomm Media Development Authority of Singapore. 2017. URL: <https://www.imda.gov.sg/news-and-events/Media-Room/archived/ida/Media-Releases/2015/home-access-programme-to-benefit-8000-low-income-households> [accessed 2020-09-11]
14. McNamee P, Murray E, Kelly MP, Bojke L, Chilcott J, Fischer A, et al. Designing and Undertaking a Health Economics Study of Digital Health Interventions. *Am J Prev Med* 2016 Nov;51(5):852-860. [doi: [10.1016/j.amepre.2016.05.007](https://doi.org/10.1016/j.amepre.2016.05.007)] [Medline: [27745685](https://pubmed.ncbi.nlm.nih.gov/27745685/)]

15. Schechter CB, Walker EA, Ortega FM, Chamany S, Silver LD. Costs and effects of a telephonic diabetes self-management support intervention using health educators. *J Diabetes Complications* 2016 Mar;30(2):300-305 [[FREE Full text](#)] [doi: [10.1016/j.jdiacomp.2015.11.017](https://doi.org/10.1016/j.jdiacomp.2015.11.017)] [Medline: [26750743](#)]
16. Jiao N, Zhu L, Chong YS, Chan WS, Luo N, Wang W, et al. Web-based versus home-based postnatal psychoeducational interventions for first-time mothers: A randomised controlled trial. *Int J Nurs Stud* 2019 Nov;99:103385. [doi: [10.1016/j.ijnurstu.2019.07.002](https://doi.org/10.1016/j.ijnurstu.2019.07.002)] [Medline: [31442783](#)]
17. Wantland DJ, Portillo CJ, Holzemer WL, Slaughter R, McGhee EM. The effectiveness of Web-based vs. non-Web-based interventions: a meta-analysis of behavioral change outcomes. *J Med Internet Res* 2004 Nov 10;6(4):e40 [[FREE Full text](#)] [doi: [10.2196/jmir.6.4.e40](https://doi.org/10.2196/jmir.6.4.e40)] [Medline: [15631964](#)]
18. Shorey S, Lau Y, Dennis C, Chan YS, Tam WWS, Chan YH. A randomized-controlled trial to examine the effectiveness of the 'Home-but not Alone' mobile-health application educational programme on parental outcomes. *J Adv Nurs* 2017 Sep;73(9):2103-2117. [doi: [10.1111/jan.13293](https://doi.org/10.1111/jan.13293)] [Medline: [28276086](#)]
19. Rollo ME, Burrows T, Vincze LJ, Harvey J, Collins CE, Hutchesson MJ. Cost evaluation of providing evidence-based dietetic services for weight management in adults: In-person versus eHealth delivery. *Nutr Diet* 2018 Feb;75(1):35-43. [doi: [10.1111/1747-0080.12335](https://doi.org/10.1111/1747-0080.12335)] [Medline: [29411491](#)]
20. Li J, Parrott S, Sweeting M, Farmer A, Ross J, Dack C, et al. Cost-Effectiveness of Facilitated Access to a Self-Management Website, Compared to Usual Care, for Patients With Type 2 Diabetes (HeLP-Diabetes): Randomized Controlled Trial. *J Med Internet Res* 2018 Jun 08;20(6):e201 [[FREE Full text](#)] [doi: [10.2196/jmir.9256](https://doi.org/10.2196/jmir.9256)] [Medline: [29884608](#)]
21. Wong BB, Chan YH, Leow MQH, Lu Y, Chong YS, Koh SSL, et al. Application of cabbage leaves compared to gel packs for mothers with breast engorgement: Randomised controlled trial. *Int J Nurs Stud* 2017 Nov;76:92-99. [doi: [10.1016/j.ijnurstu.2017.08.014](https://doi.org/10.1016/j.ijnurstu.2017.08.014)] [Medline: [28941842](#)]
22. Bandura A. *Self-efficacy: The Exercise of Control*. New York: WH Freeman; 1997.
23. Azmoude E, Jafarnejade F, Mazlom S. The predictors for maternal self-efficacy in early parenthood. *J Midwifery Reprod Health* 2015;3(2):368-376. [doi: [10.22038/JMRH.2015.4050](https://doi.org/10.22038/JMRH.2015.4050)]
24. Kohlhoff J, Barnett B. Parenting self-efficacy: links with maternal depression, infant behaviour and adult attachment. *Early Hum Dev* 2013 Apr;89(4):249-256. [doi: [10.1016/j.earlhumdev.2013.01.008](https://doi.org/10.1016/j.earlhumdev.2013.01.008)] [Medline: [23398731](#)]
25. Leahy-Warren P, McCarthy G, Corcoran P. First-time mothers: social support, maternal parental self-efficacy and postnatal depression. *J Clin Nurs* 2012 Feb;21(3-4):388-397. [doi: [10.1111/j.1365-2702.2011.03701.x](https://doi.org/10.1111/j.1365-2702.2011.03701.x)] [Medline: [21435059](#)]
26. Ngai F, Chan SWC, Ip W. Predictors and correlates of maternal role competence and satisfaction. *Nurs Res* 2010;59(3):185-193. [doi: [10.1097/NNR.0b013e3181dbb9ee](https://doi.org/10.1097/NNR.0b013e3181dbb9ee)] [Medline: [20404775](#)]
27. Barnes CR, Adamson-Macedo EN. Perceived Maternal Parenting Self-Efficacy (PMP S-E) tool: development and validation with mothers of hospitalized preterm neonates. *J Adv Nurs* 2007 Dec;60(5):550-560. [doi: [10.1111/j.1365-2648.2007.04445.x](https://doi.org/10.1111/j.1365-2648.2007.04445.x)] [Medline: [17973719](#)]
28. Shorey S, Chan SW, Chong YS, He H. Maternal parental self-efficacy in newborn care and social support needs in Singapore: a correlational study. *J Clin Nurs* 2014 Aug;23(15-16):2272-2282. [doi: [10.1111/jocn.12507](https://doi.org/10.1111/jocn.12507)] [Medline: [24372630](#)]
29. Johansson K, Darj E. What type of information do parents need after being discharged directly from the delivery ward? *Ups J Med Sci* 2004;109(3):229-238. [doi: [10.3109/2000-1967-085](https://doi.org/10.3109/2000-1967-085)] [Medline: [15508525](#)]
30. Lee AM, Lam SK, Sze Mun Lau SM, Chong CSY, Chui HW, Fong DYT. Prevalence, course, and risk factors for antenatal anxiety and depression. *Obstet Gynecol* 2007 Nov;110(5):1102-1112. [doi: [10.1097/01.AOG.0000287065.59491.70](https://doi.org/10.1097/01.AOG.0000287065.59491.70)] [Medline: [17978126](#)]
31. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. *J Psychosom Res* 2002 Feb;52(2):69-77. [doi: [10.1016/s0022-3999\(01\)00296-3](https://doi.org/10.1016/s0022-3999(01)00296-3)] [Medline: [11832252](#)]
32. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the Economic Evaluation of Health Care Programmes* (4th edn). Oxford: Oxford University Press; 2015.
33. Singapore report on registration of births and deaths. Immigration and Checkpoints Authority of Singapore. 2017. URL: <http://ghdx.healthdata.org/record/singapore-report-registration-births-and-deaths-2016> [accessed 2019-05-20]
34. Elbert NJ, van Os-Medendorp H, van Renselaar W, Ekeland AG, Hakkaart-van Roijen L, Raat H, et al. Effectiveness and cost-effectiveness of ehealth interventions in somatic diseases: a systematic review of systematic reviews and meta-analyses. *J Med Internet Res* 2014 Apr 16;16(4):e110 [[FREE Full text](#)] [doi: [10.2196/jmir.2790](https://doi.org/10.2196/jmir.2790)] [Medline: [24739471](#)]
35. Massoudi B, Holvast F, Bockting CLH, Burger H, Blanker MH. The effectiveness and cost-effectiveness of e-health interventions for depression and anxiety in primary care: A systematic review and meta-analysis. *J Affect Disord* 2019 Feb 15;245:728-743. [doi: [10.1016/j.jad.2018.11.050](https://doi.org/10.1016/j.jad.2018.11.050)] [Medline: [30447572](#)]
36. Mihalopoulos C, Vos T, Pirkis J, Carter R. The population cost-effectiveness of interventions designed to prevent childhood depression. *Pediatrics* 2012 Mar;129(3):e723-e730. [doi: [10.1542/peds.2011-1823](https://doi.org/10.1542/peds.2011-1823)] [Medline: [22312000](#)]
37. Willan AR, O'Brien BJ. Sample size and power issues in estimating incremental cost-effectiveness ratios from clinical trials data. *Health Econ* 1999 May;8(3):203-211. [doi: [10.1002/\(sici\)1099-1050\(199905\)8:3<203::aid-hec413>3.0.co;2-7](https://doi.org/10.1002/(sici)1099-1050(199905)8:3<203::aid-hec413>3.0.co;2-7)] [Medline: [10348415](#)]

38. Hall SM, Lightwood JM, Humfleet GL, Bostrom A, Reus VI, Muñoz R. Cost-effectiveness of bupropion, nortriptyline, and psychological intervention in smoking cessation. *J Behav Health Serv Res* 2005;32(4):381-392. [doi: [10.1007/BF02384199](https://doi.org/10.1007/BF02384199)] [Medline: [16215448](https://pubmed.ncbi.nlm.nih.gov/16215448/)]

Abbreviations

ICER: incremental cost-effectiveness ratio

Edited by G Eysenbach; submitted 17.11.20; peer-reviewed by J Zhu, J Li; comments to author 19.01.21; revised version received 12.05.21; accepted 18.11.21; published 11.03.22.

Please cite as:

Zheng Q, Shi L, Zhu L, Jiao N, Chong YS, Chan SWC, Chan YH, Luo N, Wang W, He H

Cost-effectiveness of Web-Based and Home-Based Postnatal Psychoeducational Interventions for First-time Mothers: Economic Evaluation Alongside Randomized Controlled Trial

J Med Internet Res 2022;24(3):e25821

URL: <https://www.jmir.org/2022/3/e25821>

doi: [10.2196/25821](https://doi.org/10.2196/25821)

PMID: [35275078](https://pubmed.ncbi.nlm.nih.gov/35275078/)

©Qishi Zheng, Luming Shi, Lixia Zhu, Nana Jiao, Yap Seng Chong, Sally Wai-Chi Chan, Yiong Huak Chan, Nan Luo, Wenru Wang, Honggu He. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org/>), 11.03.2022. 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.

Original Paper

Clinical Outcomes Among Working Adults Using the Health Integrator Smartphone App: Analyses of Prespecified Secondary Outcomes in a Randomized Controlled Trial

Stephanie Bonn¹, PhD; Gabriella Licitra¹, MD; Rino Bellocco^{2,3}, PhD; Ylva Trolle Lagerros^{1,4}, MD, PhD

¹Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

²Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

³Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy

⁴Center for Obesity, Academic Specialist Center, Stockholm Health Services, Stockholm, Sweden

Corresponding Author:

Stephanie Bonn, PhD

Clinical Epidemiology Division

Department of Medicine Solna

Karolinska Institutet

T2, Karolinska University Hospital

Stockholm, 171 76

Sweden

Phone: 46 08 51779173

Email: stephanie.bonn@ki.se

Related Article:

This is a corrected version. See correction statement: <https://www.jmir.org/2022/3/e38199/>

Abstract

Background: There is a need to find new methods that can enhance the individuals' engagement in self-care and increase compliance to a healthy lifestyle for the prevention of noncommunicable diseases and improved quality of life. Mobile health (mHealth) apps could provide large-scale, cost-efficient digital solutions to implement lifestyle change, which as a corollary may enhance quality of life.

Objective: Here we evaluate if the use of a smartphone-based self-management system, the Health Integrator app, with or without telephone counseling by a health coach, had an effect on clinical variables (secondary outcomes) of importance for noncommunicable diseases.

Methods: The study was a 3-armed parallel randomized controlled trial. Participants were randomized to a control group or to 1 of 2 intervention groups using the Health Integrator app with or without additional telephone counseling for 3 months. Clinical variables were assessed before the start of the intervention (baseline) and after 3 months. Due to the nature of the intervention, targeting lifestyle changes, participants were not blinded to their allocation. Robust linear regression with complete case analysis was performed to study the intervention effect among the intervention groups, both in the entire sample and stratifying by type of work (office worker vs bus driver) and sex.

Results: Complete data at baseline and follow-up were obtained from 205 and 191 participants, respectively. The mean age of participants was 48.3 (SD 10) years; 61.5% (126/205) were men and 52.2% (107/205) were bus drivers. Improvements were observed at follow-up among participants in the intervention arms. There was a small statistically significant effect on waist circumference ($\beta=-0.97$, 95% CI -1.84 to -0.10) in the group receiving the app and additional coach support compared to the control group, but no other statistically significant differences were seen. However, participants receiving only the app had statistically significantly lower BMI ($\beta=-0.35$, 95% CI -0.61 to -0.09), body weight ($\beta=-1.08$, 95% CI -1.92 to -0.26), waist circumference ($\beta=-1.35$, 95% CI -2.24 to -0.45), and body fat percentage ($\beta=-0.83$, 95% CI -1.65 to -0.02) at follow-up compared to the controls. There was a statistically significant difference in systolic blood pressure between the two intervention groups at follow-up ($\beta=-3.74$, 95% CI -7.32 to -0.16); no other statistically significant differences in outcome variables were seen.

Conclusions: Participants randomized to use the Health Integrator smartphone app showed small but statistically significant differences in body weight, BMI, waist circumference, and body fat percentage compared to controls after a 3-month intervention. The effect of additional coaching together with use of the app is unclear.

Trial Registration: ClinicalTrials.gov NCT03579342; <https://clinicaltrials.gov/ct2/show/NCT03579342>

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-019-6595-6

(*J Med Internet Res* 2022;24(3):e24725) doi:[10.2196/24725](https://doi.org/10.2196/24725)

KEYWORDS

adults; body composition; exercise; HbA1c; healthy lifestyle; metabolic health; mobile app; randomized controlled trial; smartphone

Introduction

On the World Health Organization (WHO) list of the top 10 causes of death for 2020, three of the four top causes, ischemic heart disease, stroke, and chronic obstructive pulmonary disease, are all noncommunicable diseases strongly linked to lifestyle [1]. The etiology of such diseases includes lifestyle-related risk factors such as physical inactivity, unhealthy diet, and tobacco use. WHO estimates that almost 90% of premature deaths due to noncommunicable diseases occur in low- and middle-income countries.

Mobile health (mHealth), defined as “medical and public health practice supported by mobile devices, such as mobile phones, patient monitoring devices, personal digital assistants, and other wireless devices” is one approach to take global action for the prevention and control of noncommunicable diseases [2]. WHO has acknowledged the potential benefit of mHealth for public health and called for evidence of effectiveness [3].

The rapid evolution of information technology has made smartphones an integral part of daily life. Worldwide, 48% of people own a smartphone [4]; in Sweden, this number is 92% [5]. Furthermore, ownership and use are independent of socioeconomic status. This makes Sweden a suitable country to evaluate intervention research using smartphone technology, which later can be disseminated to other populations.

A number of studies have been conducted on the use of smartphone technology to implement and support lifestyle changes among persons with noncommunicable diseases such as diabetes [6,7]. A systematic review of the literature found that the most common topics for health promotion programs using mobile apps were diet and physical activity, and across the studies included, app users were more successful in terms of health outcomes [8]. A meta-analysis including more than 6000 participants in app-based mobile interventions for improving nutrition behaviors and nutrition-related health outcomes did not find any evidence that additional intervention components or number of behaviour change techniques included in the intervention mattered [9]. Contrasting these results, authors of another systematic review concluded that multicomponent interventions that combined the smartphone app with some other kind of intervention appeared more effective than stand-alone app interventions where the app was the only intervention component [10]. The use of additional health coaches complementing an app has been evaluated in lifestyle interventions targeting for example persons with diabetes [11-13]. While health coaches can help to set realistic

goals and encourage when motivation fails, they are also less scalable. Whereas some studies have shown favorable results in the smartphone and coach-assisted groups compared to control groups [11,13], others could not find a significant difference between groups [12]. Clearly, a smartphone app available at any time is more convenient compared to scheduled face-to-face or group interventions, but it has not been convincingly documented if multicomponent smartphone-assisted lifestyle interventions, with or without health coaches, are effective in changing lifestyle behaviors.

Based on evidence from previous research on smartphone-assisted lifestyle interventions, we developed and built a new digital platform and an adjacent smartphone app for lifestyle change called the Health Integrator as part of a European collaboration granted by the European Institute of Innovation and Technology (EIT).

The Health Integrator core system consists of a home page, a smartphone app, and underlying supporting technical services and infrastructure. The platform was developed to be a collaboration tool for individuals in need of lifestyle change and their health coaches as a means to provide a personalized health improvement program. The program offers a variety of public, private, and community services for behavior change in different domains such as smoking, alcohol, physical activity, diet, stress, and sleep. These offers can be accessed through the Health Integrator smartphone app for those in need of lifestyle change who have a personal user account.

The aim of this study was to evaluate if use of the Health Integrator, offering a multicomponent lifestyle intervention for 3 months, with or without additional counseling by a health coach, had an effect on clinical variables (secondary outcomes) of importance for noncommunicable diseases such as BMI, waist circumference, and blood pressure in working adults and, more specifically, in office workers and bus drivers.

Methods

Study Design

The design of the Health Integrator study has been described in detail elsewhere [14]. In brief, the study was a 3-armed parallel randomized controlled trial with allocation 1:1:1, to one of two intervention arms or a control group. The active intervention was 3 months. Assessment of lifestyle behavioral factors and clinical variables was done before the start of the intervention (baseline) and after completed intervention at 3-month follow-up. Power was calculated based on the main

outcome, change in general health perception [14], and the intervention was a priori planned to include at least 63 participants in each study arm. The intervention was described according to the CONSORT-EHEALTH checklist [15].

Participants

Both men and women were eligible for participation and inclusion criteria were being aged 18 years or older, understanding Swedish well enough to understand the study aims, providing informed consent for participation, and having access and ability to use a smartphone.

Study participants were recruited from 4 companies: 2 companies with white-collar employees (ie, office workers) and 2 companies with blue-collar employees, (ie, bus drivers). After discussions with the offices of the human relations at the different companies, recruitment to the study was conducted in 2 different ways according to the wishes of the companies. White-collar employees got an email from the office of human relations with information about the study. Those who were interested in participating responded by email, and research personnel then received their email addresses. Bus drivers, on the other hand, whose email addresses were unknown to the employer, were informed by study personnel at the bus garages and asked in person to provide their private email. Thereafter, study personnel emailed detailed information about the study and a link to access the web-based baseline questionnaire to all employees interested in the study.

Data Collection

After having completed the baseline questionnaire, all respondents were provided with a link to the Health Integrator system. They were asked to answer additional questions creating a health profile in the system and could thereafter schedule a time for the baseline meeting with the health coach. At this point, participants had not been randomized and the results from the health profile questions were not yet visible to the participant in the Health Integrator system.

All study participants meet with study personnel at baseline and again after 3-month follow-up. All participants received an accelerometer to assess physical activity during 7 consecutive days during the baseline visit, and body composition and blood pressure were measured on both visits. At baseline, results from the health profile were also unlocked for participants in the 2 intervention groups but not for participants in the control group. All participants also received a referral for analysis of glycated hemoglobin A_{1c} (HbA_{1c}) and serum lipids. At the 3-month follow-up, only participants with an HbA_{1c} or serum lipids outside clinical reference values were given a second referral. Before the 3-month follow-up meeting, participants responded to the web-based questionnaire and the additional health profile questions again. Participants also responded to a final web-based questionnaire after 6 months.

Randomization

Participants were randomized to 1 of 3 groups: intervention group A receiving the Health Integrator smartphone app and additional coach support, intervention group B receiving the Health Integrator smartphone app without additional coach

support, and control group C did not receive the Health Integrator smartphone app or any coach support. Participants in the control group received access to the app after the active intervention had been completed at 3-month follow-up. Randomization was done in blocks of 6 by company and sex using a random allocation list generated by the first author using Stata (version 15.1, StataCorp LLC). Each new participant included in the study was continuously assigned to the next available randomly allocated group on the computer-generated list by the first author. Participants were randomized before the baseline meeting and informed about their allocation when meeting with study personnel at baseline measurements. Study personnel were not blinded to the allocation during the meeting but did not reveal the allocation to the participant until after performing baseline assessments. Due to the nature of the intervention, participants were not blinded to their allocation.

Intervention

During the baseline meeting, participants that had been randomized to any of the 2 intervention groups (with and without additional coach support) downloaded the Health Integrator smartphone app. The app was compatible with both Android (version 4.1 and higher) and iOS (version 8 and higher). User satisfaction with the app was assessed at the end of the trial [16].

Results from the health profile were discussed with the health coach at the baseline meeting. The participant and the health coach decided on which lifestyle behavior to target based on the health profile. Thereby, each participant received a personalized intervention, customized based on their needs and goals. The intervention could target any of the following 6 areas: diet, physical activity, sleeping habits, stress, alcohol, or tobacco use.

Within the Health Integrator system, a number of different offers related to the different intervention areas were available. For example, a participant aiming to increase physical activity levels could choose between offers including, for example, other smartphone apps developed to promote physical activity specifically (eg, Runkeeper), to get a wrist support band to facilitate rehabilitation, receive a training pass at a local gym, or set the goal to take part in a race and have the fee paid for. There were in total 37 offers. The offers were free of charge for the participant.

Related to the target area, the health coach together with the participant identified and agreed on a suitable weekly goal, which was entered into the Health Integrator. This could, for example, be number of sessions of physical activity per week. To keep on track during the 3 months of active intervention, participants recorded if the weekly goal was met using an emoticon scale (ie, smiley faces) or by marking the number of days that the goal was met during the week. A reminder to record progression was sent out every Sunday at 9:20 PM. Those participants randomized to receive the Health Integrator smartphone app and additional health coach support had a scheduled telephone appointment every 4 weeks with the health coach. During these 30-minute sessions, participants were given personalized guidance and encouragement to support the

lifestyle change in question. During these sessions, it was also possible to update or change the weekly goal.

Control Group

During the active intervention for the intervention groups, the control group participants were not given any lifestyle recommendations nor were they aware of which offers were included in the app. Results from the baseline health profile were shown and discussed with the health coach at the 3-month follow-up (ie, after the active intervention). At that time, participants in the control group were offered to download the Health Integrator app and were given access to the same offers as the intervention groups.

Outcome Measures

Self-reported data on civil status (having a partner vs not having a partner), smoking (yes/no), snuff use (yes/no), physical activity (<150 min/wk vs 150-300 min/wk vs ≥ 300 min/wk), and treatment for diabetes (yes/no), hypertension (yes/no), high serum lipids (yes/no), and diabetes risk using the Finnish Diabetes Risk Score (FINDRISC) were retrieved from the extensive web-based baseline questionnaire [17]. See [Multimedia Appendix 1](#) for a complete list of study assessments in the trial.

Weight (kg), waist circumference (cm), body fat (percentage), and blood pressure (mm Hg) were measured by study personnel at baseline and follow-up after 3 months. Height (cm) was self-reported at baseline. BMI (kg/m^2) was calculated based on measured weight and reported height. Body weight, waist circumference, and body fat percentage were analyzed separately for women and men.

At baseline, all participants received a referral for analysis of HbA_{1c} (mmol/mol), total cholesterol (mmol/L), apolipoprotein A1 (g/L), and apolipoprotein B (g/L).

Ethical Considerations

The study was approved by the regional ethical review board in Stockholm, Sweden (2018/411-31 and 2018/1038-32). The trial was registered at ClinicalTrials.gov [NCT03579342]. Eligible participants were required to give their informed consent prior to responding to the baseline questionnaire. After reading an introductory screen displaying information about the study, participants were required to consent to participate in order to continue to the questionnaire. At the baseline meeting, participants also gave their written informed consent.

Statistical Analysis

Baseline demographics and clinical characteristics were summarized using descriptive statistics. Continuous variables are shown as mean and standard deviation and categorical variables as frequency and percentage. Results are shown for

all study participants and by study group (intervention group A, B, and control group C).

We performed robust linear regression (robreg in Stata), 1-way robust analysis of covariance [18], to analyze how the outcome variables at follow-up differed among treatment groups (intervention with or without counseling and control). We adjusted for baseline measurements of BMI, weight, waist circumference, body fat percentage, and systolic and diastolic blood pressure, and since randomization was done by company and sex, we also adjusted for those variables.

The modern robust approach finds parameter estimates that are less sensitive to outliers and influential data, while at the same time retaining statistical efficiency by minimizing a different target function, which give less weight to individuals with large residuals. The regression coefficients (β) with 95% confidence intervals were estimated based on an iteratively reweighted ordinary least squares algorithm.

We also conducted prespecified subgroup analyses by company where the effect of the intervention between intervention groups and control group (intervention group A vs control group C, intervention group B vs control group C, and intervention group A vs intervention group B) was analyzed.

Missing data was minimal with 99% of participants having complete data on body weight and BMI and 93% having complete data on blood pressure, waist circumference, and body fat percentage. Therefore, all statistical analyses were based on complete cases [19]. The statistical significance was set at $P < .05$. All analyses were completed using Stata.

Results

In total, 209 participants were recruited to the study. Among these, 4 did not complete baseline measurements and were excluded from all analysis. At the 3-month follow-up, 191 participants had complete data. At baseline, one participant was missing data on weight (1/205, 0.5%), and waist circumference, body fat percentage, and blood pressure were missing from 14 participants (14/205, 6.8%).

Baseline characteristics of all participants divided into the different intervention groups are shown in [Table 1](#). There were 61.5% (126/205) men in the study, 47.8% (98/205) of the participants were office workers, and 52.2% (107/205) were bus drivers. The mean age was 48.3 years. Mean BMI of 27.1 kg/m^2 indicated overweight, which was in line with a mean waist circumference of 83.1 cm among women and 97.7 cm among men. A waist circumference above 80 cm for women and 94 cm for men is associated with an increased disease risk according to WHO [20]. The mean FINDRISC of 8.6 indicated a somewhat increased risk of developing diabetes type 2 (1/25 individuals compared to 1/100 within 10 years) [21].

Table 1. Characteristics of study participants in the Health Integrator study.

	All (n=205)	Intervention group A (n=70)	Intervention group B (n=68)	Intervention group C (n=67)
Age (years), mean (SD)	48.3 (10.0)	48.7 (10.7)	47.9 (10.5)	48.3 (8.8)
Sex, n (%)				
Women	79 (38.5)	27 (38.6)	27 (39.7)	25 (37.3)
Men	126 (61.5)	43 (61.4)	41 (60.3)	42 (62.7)
BMI (kg/m ²), mean (SD)	27.1 (4.6)	27.6 (4.6)	26.7 (4.7)	27.2 (4.6)
Body weight (kg), mean (SD)				
Women	72.8 (14.0)	75.9 (16.6)	69.8 (14.0)	72.6 (10.5)
Men	88.6 (15.8)	87.5 (12.9)	88.4 (16.6)	90.0 (17.9)
Waist circumference (cm), mean (SD)				
Women	83.1 (12.3)	82.3 (10.2)	81.2 (12.2)	85.7 (14.1)
Men	97.7 (12.9)	97.0 (11.8)	97.1 (13.7)	98.9 (13.4)
Body fat (%), mean (SD)				
Women	35.6 (7.8)	37.0 (8.5)	34.5 (7.0)	35.4 (8.0)
Men	26.2 (7.1)	25.6 (7.0)	26.4 (7.8)	26.4 (6.6)
Blood pressure (mm Hg), mean (SD)				
Systolic	129 (17)	129 (14)	129 (21)	130 (14)
Diastolic	82 (10)	81 (8)	82 (12)	82 (8)
Type of work, n (%)				
Office worker	98 (47.8)	31 (44.3)	34 (50.0)	33 (49.3)
Bus driver	107 (52.2)	39 (55.7)	34 (50.0)	34 (50.7)
Civil status, n (%)				
Partner	163 (79.5)	56 (80.0)	59 (86.8)	48 (71.6)
No partner	36 (17.6)	10 (14.3)	9 (13.2)	17 (25.4)
Smoking, n (%)				
Yes	18 (8.8)	4 (5.7)	4 (5.9)	10 (14.9)
No	187 (91.2)	66 (94.3)	64 (94.1)	57 (85.1)
Snuff use, n (%)				
Yes	21 (10.2)	9 (12.9)	3 (4.4)	9 (13.4)
No	184 (89.8)	61 (87.1)	65 (95.6)	58 (86.6)
Weekly physical activity (min), n (%)				
≥300	64 (31.2)	20 (28.6)	25 (36.8)	19 (28.4)
150-300	54 (26.3)	19 (27.1)	19 (27.9)	16 (23.9)
<150	84 (41.0)	29 (41.4)	24 (35.3)	31 (46.37)
History of treatment, n (%)				
Diabetes (yes)	8 (3.9)	3 (4.3)	2 (2.9)	3 (4.5)
High blood pressure (yes)	26 (12.7)	10 (14.3)	13 (19.1)	3 (4.5)
High serum lipid levels (yes)	9 (4.4)	5 (7.1)	2 (2.9)	2 (3.0)
FINDRISC ^a , mean (SD)	8.6 (5.3)	9.1 (5.7)	8.2 (5.2)	8.3 (5.0)
HbA _{1c} ^b (mmol/mol), mean (SD)	36.9 (6.1)	37.1 (7.4)	36.8 (5.4)	36.7 (5.3)
HbA _{1c} (%), mean (SD)	5.5 (2.7)	5.5 (2.8)	5.5 (2.6)	5.5 (2.6)

	All (n=205)	Intervention group A (n=70)	Intervention group B (n=68)	Intervention group C (n=67)
Total cholesterol (mmol/L), mean (SD)	4.85 (1.0)	4.8 (1.0)	4.9 (1.0)	4.9 (0.9)
Apolipoprotein A1 (g/L), mean (SD)	1.43 (0.3)	1.44 (0.3)	1.42 (0.3)	1.44 (0.3)
Apolipoprotein B (g/L), mean (SD)	0.96 (0.3)	0.95 (0.3)	0.97 (0.3)	0.96 (0.2)

^aFINDRISC: Finnish Diabetes Risk Score.

^bHbA_{1c}: hemoglobin A_{1c}.

Results of intervention effects are shown in Table 2. There was a statistically significant intervention effect with intervention group A having a mean waist circumference of approximately 1 cm less compared to the control group at follow-up ($\beta=-0.97$, 95% CI -1.84 to -0.10). No other statistically significant differences were seen, although point estimates all favor the intervention group. Comparing intervention group B to the control group, a lower BMI, body weight, waist circumference and body fat percentage were seen at follow-up (BMI: $\beta=-0.35$,

95% CI -0.61 to -0.09 ; body weight: $\beta=-1.08$, 95% CI -1.92 to -0.26 ; waist circumference: $\beta=-1.35$, 95% CI -2.24 to -0.45 ; body fat percentage $\beta=-0.83$, 95% CI -1.65 to -0.02). We found no statistically significant difference between group B and the control group with regards to blood pressure. When comparing intervention group A to intervention group B, there was a statistically significant difference in systolic blood pressure ($\beta=-3.74$, 95% CI -7.32 to -0.16). No other statistically significant differences were seen when comparing the groups.

Table 2. Comparison of intervention effect between intervention groups and control group among all study participants, complete case analysis using robust regression, the Health Integrator study^a.

	Intervention group A vs control group C, β (95% CI)	Intervention group B vs control group C, β (95% CI)	Intervention group A vs intervention group B, β (95% CI)
BMI (kg/m ²)	-0.22 (-0.47 to 0.03)	-0.35 (-0.61 to -0.09)	0.13 (-0.13 to 0.38)
Weight (kg)	-0.69 (-1.45 to 0.07)	-1.08 (-1.92 to -0.26)	0.40 (-0.40 to 1.19)
Waist circumference (cm)	-0.97 (-1.84 to -0.10)	-1.35 (-2.24 to -0.45)	0.37 (-0.54 to 1.28)
Body fat (%)	-0.71 (-1.44 to 0.02)	-0.83 (-1.65 to -0.02)	0.12 (-0.65 to 0.88)
SBP ^b (mm Hg)	-1.24 (-5.01 to 2.53)	2.50 (-1.56 to 6.56)	-3.74 (-7.32 to -0.16)
DBP ^c (mm Hg)	-0.25 (-2.85 to 2.35)	1.63 (-0.86 to 4.12)	-1.88 (-4.17 to 0.40)

^aAdjusted for sex, company, and baseline levels of BMI, weight, waist circumference, body fat percentage, and systolic and diastolic blood pressure.

^bSBP: systolic blood pressure.

^cDBP: diastolic blood pressure.

Stratified analyses, comparing intervention effects among office workers and bus drivers, are shown in Multimedia Appendix 2. Among office workers, female office workers in intervention group B had a statistically significantly lower body weight and waist circumference compared to women in the control group (body weight: $\beta=-1.66$, 95% CI -3.18 to -0.15 ; waist circumference: $\beta=-1.82$, 95% CI -3.11 to -0.52), and male office workers had a lower waist circumference ($\beta=-2.35$, 95% CI -3.93 to -0.77). No other differences were seen among office workers when comparing intervention group A or intervention group B with the control group, or when comparing intervention groups A against intervention group B, except that male office workers in intervention group A had a statistically significantly higher waist circumference ($\beta=2.74$, 95% CI 0.87 to 4.60) and lower body fat percentage ($\beta=-2.92$, 95% CI -0.89 to -0.95) compared to intervention group B after the intervention.

Among bus drivers, BMI at follow-up was lower in both intervention group A ($\beta=-0.38$, 95% CI -0.67 to -0.10) and intervention group B ($\beta=-0.45$, 95% CI -0.77 to -0.12) compared to the control group. Among male bus drivers in both intervention group A and intervention group B, body weight

and waist circumference were statistically significantly lower at follow-up compared to the control group (Group A vs C: body weight: $\beta=-1.13$, 95% CI -2.14 to -0.12 , waist circumference: $\beta=-1.73$, 95% CI -3.37 to -0.08 , and group B vs C: body weight: $\beta=-1.38$, 95% CI -2.54 to -0.22 ; waist circumference: $\beta=-1.65$, 95% CI -3.05 to -0.26). No other statistically significant differences were seen among bus drivers when comparing intervention group A or intervention group B with the control group, or when comparing intervention group A to intervention group B.

Discussion

Principal Findings

Our results show that using an mHealth solution such as the Health Integrator smartphone app, focusing on an array of lifestyle habits, may have beneficial effects on health markers of clinical importance, such as weight, waist circumference, and body fat percentage, although the effect sizes were small in this 3-month trial. For overweight persons or persons with obesity, the mean initial 3-month weight loss for the entire

sample is too small to be clinically meaningful, but from a public health perspective, small benefits are also appreciable [22].

In our study, those randomized to the group that only met the health coach at baseline and at the end of the study were found to have statistically significant improvements in more clinical variables than those randomized to be in contact with the health coach every 4 weeks. If one group's intervention is nested within the other (B is within A), then the effect should be expected to be present in group A if it is present in group B, but this was only seen for waist circumference. The reasons for the inconsistent finding raises concerns, and with multiple secondary outcomes, there is a risk of statistical significance by chance.

Comparison With Previous Findings

The exponential growth of mobile communication has been coupled with an increase in trials using mHealth. Many previous studies have been feasibility studies with relatively few participants [13,23,24]. However, some mHealth studies are comparable in size and scope to ours. The multicomponent mHealth intervention by Spring et al [25] targeting diet and physical activity in 212 North Americans with low fruit and vegetable and high saturated fat intake, low moderate to vigorous activity, and high screen time is similar to our intervention. They reported statistically significant improvements in the intervention groups compared to the controls in terms of both diet and physical activity. They too included regular calls from a trained paraprofessional in addition to using an app. Controls, however, received an app targeting sleep and stress instead of diet and physical activity, while controls in our study received standard care. Intriguingly, having access to a health coach calling twice during our intervention did not lead to better effect of the intervention in terms of clinical outcomes. A similarly designed 3-armed Norwegian mHealth intervention with and without telephone counseling targeting self-management and lifestyle change for persons with type 2 diabetes did not find a better intervention effect among those receiving telephone counseling [12]. Potentially, studies about the mechanisms of change (ie, the processes that operate in behavior change interventions) could be useful to improve delivery of future behavior change interventions.

While weight loss is reasonable to expect in interventions targeting diet, physical activity, or even alcohol habits, previous studies have shown that stress and poor sleep can be important barriers for losing weight [26]. Acute stress tends to increase the preference for sweet food and eating in the absence of hunger, leading to an increased energy intake [27]. This poses the intriguing possibility that the commonly poor outcome of many weight loss programs could benefit from interventions targeting other lifestyle habits that the participants themselves identify as important, other than diet.

Limitations and Strengths

A limitation to our trial might be the fact that participants were not blinded to their group allocation during the trial. Blinding in randomized trials prevents bias due to, for example, the use of co-interventions or biased ascertainment of outcomes. It is difficult to blind participants in lifestyle interventions. While Spring et al [25] gave the controls an app focusing on sleep and

stress instead of the active intervention targeting physical activity and diet, this was not an available option in our study, since we also targeted sleep and stress. Potentially, the controls could have been given a sham lifestyle regimen deemed to be ineffective. For ethical reasons, giving everyone an equal opportunity to improve their lifestyle, we chose to use wait-list controls (ie, after 3 months); when the active intervention was finished, the controls met the health coaches, discussed their health profile, and received access to all offers at the Health Integrator. Although the control group participants were not aware of which offers were included in the app during the first 3 months of the trial, this did not preclude that some controls could have found other ways to improve their lifestyle already during the trial, which would have led to smaller effect sizes.

Blinding of personnel may prevent biased ascertainment of outcome, if the investigator is tempted to look more carefully for certain outcomes in one or the other group. Our laboratory personnel analyzing the blood samples taken were blinded, but other study personnel were not. Nonetheless, clinical variables such as weight, waist circumference, and blood pressure are objective measures and should be less prone to be biased by unblinded staff.

Social desirability, a natural reaction to defend one's social image when using self-reported measures, leading to a tendency to overestimate desirable behaviors and underestimate undesirable behaviors [28] may have led to overestimation of, for example, baseline characteristics such as physical activity. Among our participants, 57.5% self-reported meeting the general recommendation of at least 150 minutes of moderate-to-vigorous physical activity per week at baseline. When a population sample of more than a thousand Swedish adults aged 50 to 64 years had their physical activity objectively measured with a hip-worn accelerometer for a week, even more (72.5%) reached the recommendation [29]. It should be kept in mind, however, that wearing an accelerometer may itself affect behavior, the so-called Hawthorne effect. Since all participants in the study also wore an accelerometer at baseline and follow-up, this could have affected behavior among the participants and among the controls. This may be a limitation when evaluating the effect of the intervention, since this potentially led to less difference between groups. It should be noted, however, that the accelerometers did not display any recordings of activity, and none of the participants were informed about their accelerometer data.

A common limitation in intervention studies is dropout and attrition of study participants. In a previous Swedish study evaluating an internet-based weight loss program, only 19.4% (4440/22,860) logged in at least twice during the first 3 months and at least twice during the last 2 months [30]. Nevertheless, compliance in our study was high both in the intervention groups and in the control group. Over 90% of participants had complete data at 3-month follow-up.

While most interventions solely focus on one health behavior such as physical activity [31] or stress [32] or target a specific patient group, such as persons with diabetes [33], pregnant women [34], or patients postsurgery [35], our study had a person-centered approach. The Health Integrator aimed to find

the specific lifestyle behavior that the person in question could improve and was interested in improving. This may also have been key to the high compliance in our study. Nevertheless, the fact that participants worked on different types of interventions may have led to power issues in analyses of secondary outcomes. Future researchers may want to include more participants and have longer follow-up.

Another strength of our study is the objective measurement of clinical outcomes, such as waist circumference, body fat percentage, blood pressure, HbA_{1c}, and serum lipids. However, the lack of data on HbA_{1c} and serum lipids (ie, total cholesterol, apolipoprotein A1 and apolipoprotein B) at follow-up is a limitation. There were few participants with baseline values outside clinical references and, following clinical practice, only those with pathological values outside the reference were referred for a second blood sampling at the follow-up assessment. Therefore, it was not possible to evaluate effects of the intervention in these markers. Besides, it is unlikely that we would have detected a difference from baseline to follow-up given that the majority of the participants had values within the reference at baseline. However, according to a systematic review including 23 mHealth interventions targeting lifestyle in persons with diabetes, the effect on HbA_{1c} was statistically significant [7].

Generalizability

Our randomized study included both men and women as well as employees from different types of professions (ie, bus drivers and office workers), potentially representing different socioeconomic groups in society. However, since employment was a prerequisite for participation in our study, this may have created a selection of healthier people than the general population. This selection, which in cohort studies is known as the healthy worker effect, is not an issue in terms of internal validity. However, in terms of external validity, this might translate to an underestimation of the potential efficacy of the

intervention. Targeting high-risk unhealthy populations is a well-known way to achieve efficiency and return on investment of an intervention, since there is a greater available risk to be reduced [22].

The recruitment process was slightly different for office workers, who initially got an email from the office of human relations with information about the study, while the study personnel on site in the bus garages orally informed the bus drivers. This could have led to differences in which individuals decided to take part in the study. For example, we know that some bus drivers signed up to get more information about the study together with a colleague. Those participants may have been more compelled to pursue lifestyle changes when they felt supported and connected with colleagues doing the same than if they had signed up alone, which may be more likely when a person receives an email. Peer support during lifestyle intervention has been shown to promote health behavior change [36,37].

The Health Integrator smartphone app was available for both iOS and Android devices. Hence, most smartphone users were able to participate in the study. Nevertheless, the fact that you needed to have access and the ability to handle a smartphone may be a limitation. However, smartphone use in Sweden is widespread, with 92% of the population over age 16 years owning their own smartphone [38]. Thus, the inclusion criteria of having a smartphone is likely not a major limitation in our study setting.

Conclusion

To conclude, among study participants using only our app that targets the individual's specific need of lifestyle change, we found small statistically significant differences in body weight, BMI, body fat percentage, and waist circumference after a 3-month intervention when compared to the control group. The effect of additional coaching together with use of the app was unclear.

Acknowledgments

We would like to acknowledge our health coaches Tove Högman and Carolina Wierbecki for their hard work and dedication. Furthermore, we would like to thank Ellen Franzén, Olov Ståhl, and Per Hansson for their thorough work in developing the Health Integrator system; Petra Sommarlund for excellent European Institute of Innovation and Technology (EIT) coordination; and Fredrik Söder for his devotion to find collaborators and service offers for the Health Integrator. The Health Integrator was developed as is a part of the European collaboration "The health movement – for patient empowerment." This trial was one of the work packages (work package ID A1803) supported by grant 17088 from the EIT.

Editorial notice: This randomized study was only retrospectively registered as the authors were given the opportunity to begin data collection earlier than planned. As such, the trial was retrospectively registered three weeks after the start of the study. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as retrospective registration does not prevent authors from changing their outcome measures retrospectively.

Authors' Contributions

YTL designed and is responsible for the study. SB performed the data collection. GL performed the analyses, and results were critically reviewed by SB, RB, and YTL. The initial draft of the manuscript was prepared by SB and YTL. All authors have critically reviewed and edited the manuscript and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Study assessments in the Health Integrator randomized controlled trial.

[[DOCX File, 18 KB - jmir_v24i3e24725_app1.docx](#)]

Multimedia Appendix 2

Comparison of intervention effect between intervention groups and control group stratified by type of work; office workers and bus drivers, complete case analysis using robust regression.

[[DOCX File, 18 KB - jmir_v24i3e24725_app2.docx](#)]

Multimedia Appendix 3

CONSORT eHealth Checklist (V1.6.1).

[[PDF File \(Adobe PDF File\), 1646 KB - jmir_v24i3e24725_app3.pdf](#)]

References

1. Global Health Observatory data. World Health Organization. URL: https://www.who.int/gho/mortality_burden_disease/causes_death/top_10/en/ [accessed 2020-09-08]
2. Global action plan for the prevention and control of noncommunicable diseases. World Health Organization. 2013. URL: <https://apps.who.int/iris/rest/bitstreams/442296/retrieve> [accessed 2022-02-22]
3. mHealth: New horizons for health through mobile technologies: second global survey on eHealth. World Health Organization. 2011. URL: https://apps.who.int/iris/bitstream/handle/10665/44607/9789241564250_eng.pdf?sequence=1&isAllowed=y [accessed 2022-02-22]
4. How many smartphones are in the world?. 2020. URL: <https://www.bankmycell.com/blog/how-many-phones-are-in-the-world> [accessed 2020-09-08]
5. The Swedes and Internet 2018 (Svenskarna och internet 2018). URL: <https://www.iis.se/fakta/svenskarna-och-internet-2018/> [accessed 2020-09-08]
6. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. *J Med Internet Res* 2018 May 04;20(5):e162 [FREE Full text] [doi: [10.2196/jmir.9751](https://doi.org/10.2196/jmir.9751)] [Medline: [29728346](https://pubmed.ncbi.nlm.nih.gov/29728346/)]
7. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: [10.2196/12297](https://doi.org/10.2196/12297)] [Medline: [30664494](https://pubmed.ncbi.nlm.nih.gov/30664494/)]
8. Lee M, Lee H, Kim Y, Kim J, Cho M, Jang J, et al. Mobile app-based health promotion programs: a systematic review of the literature. *Int J Environ Res Public Health* 2018 Dec 13;15(12):1. [doi: [10.3390/ijerph15122838](https://doi.org/10.3390/ijerph15122838)] [Medline: [30551555](https://pubmed.ncbi.nlm.nih.gov/30551555/)]
9. Villinger K, Wahl DR, Boeing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: a systematic review and meta-analysis. *Obes Rev* 2019 Oct;20(10):1465-1484. [doi: [10.1111/obr.12903](https://doi.org/10.1111/obr.12903)] [Medline: [31353783](https://pubmed.ncbi.nlm.nih.gov/31353783/)]
10. 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](https://doi.org/10.1186/s12966-016-0454-y)] [Medline: [27927218](https://pubmed.ncbi.nlm.nih.gov/27927218/)]
11. 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;17(10):e224 [FREE Full text] [doi: [10.2196/jmir.4871](https://doi.org/10.2196/jmir.4871)] [Medline: [26441467](https://pubmed.ncbi.nlm.nih.gov/26441467/)]
12. Holmen H, Torbjørnsen A, Wahl AK, Jennum 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;2(4):e57 [FREE Full text] [doi: [10.2196/mhealth.3882](https://doi.org/10.2196/mhealth.3882)] [Medline: [25499872](https://pubmed.ncbi.nlm.nih.gov/25499872/)]
13. Wang J, Cai C, Padhye N, Orlander P, Zare M. A behavioral lifestyle intervention enhanced with multiple-behavior self-monitoring using mobile and connected tools for underserved individuals with type 2 diabetes and comorbid overweight or obesity: pilot comparative effectiveness trial. *JMIR Mhealth Uhealth* 2018 Apr 10;6(4):e92 [FREE Full text] [doi: [10.2196/mhealth.4478](https://doi.org/10.2196/mhealth.4478)] [Medline: [29636320](https://pubmed.ncbi.nlm.nih.gov/29636320/)]
14. Bonn SE, Löf M, Östenson C, Trolle Lagerros Y. App-technology to improve lifestyle behaviors among working adults: the Health Integrator study, a randomized controlled trial. *BMC Public Health* 2019 Mar 07;19(1):273 [FREE Full text] [doi: [10.1186/s12889-019-6595-6](https://doi.org/10.1186/s12889-019-6595-6)] [Medline: [30845949](https://pubmed.ncbi.nlm.nih.gov/30845949/)]
15. Eysenbach G. CONSORT-EHEALTH: implementation of a checklist for authors and editors to improve reporting of web-based and mobile randomized controlled trials. *Stud Health Technol Inform* 2013;192:657-661. [Medline: [23920638](https://pubmed.ncbi.nlm.nih.gov/23920638/)]

16. Melin J, Bonn SE, Pendrill L, Trolle Lagerros Y. A questionnaire for assessing user satisfaction with mobile health apps: development using Rasch measurement theory. *JMIR Mhealth Uhealth* 2020 May 26;8(5):e15909 [FREE Full text] [doi: [10.2196/15909](https://doi.org/10.2196/15909)] [Medline: [32452817](https://pubmed.ncbi.nlm.nih.gov/32452817/)]
17. Martin E, Ruf E, Landgraf R, Hauner H, Weinauer F, Martin S. FINDRISK questionnaire combined with HbA1c testing as a potential screening strategy for undiagnosed diabetes in a healthy population. *Horm Metab Res* 2011 Oct;43(11):782-787. [doi: [10.1055/s-0031-1286333](https://doi.org/10.1055/s-0031-1286333)] [Medline: [22009373](https://pubmed.ncbi.nlm.nih.gov/22009373/)]
18. Verardi V, Croux C. Robust Regression in Stata. *Stata J* 2009 Sep 01;9(3):439-453. [doi: [10.1177/1536867x0900900306](https://doi.org/10.1177/1536867x0900900306)]
19. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials: a practical guide with flowcharts. *BMC Med Res Methodol* 2017 Dec 06;17(1):162 [FREE Full text] [doi: [10.1186/s12874-017-0442-1](https://doi.org/10.1186/s12874-017-0442-1)] [Medline: [29207961](https://pubmed.ncbi.nlm.nih.gov/29207961/)]
20. Waist circumference and waist-hip ratio: report of a WHO expert consultation. World Health Organization. 2008. URL: https://www.who.int/nutrition/publications/obesity/WHO_report_waistcircumference_and_waisthip_ratio/en/ [accessed 2022-02-22]
21. Lindström J, Absetz P, Hemiö K, Peltomäki P, Peltonen M. Reducing the risk of type 2 diabetes with nutrition and physical activity: efficacy and implementation of lifestyle interventions in Finland. *Public Health Nutr* 2010 Jun;13(6A):993-999. [doi: [10.1017/S1368980010000960](https://doi.org/10.1017/S1368980010000960)] [Medline: [20513271](https://pubmed.ncbi.nlm.nih.gov/20513271/)]
22. Fisher EB, Fitzgibbon ML, Glasgow RE, Haire-Joshu D, Hayman LL, Kaplan RM, et al. Behavior matters. *Am J Prev Med* 2011 May;40(5):e15-e30 [FREE Full text] [doi: [10.1016/j.amepre.2010.12.031](https://doi.org/10.1016/j.amepre.2010.12.031)] [Medline: [21496745](https://pubmed.ncbi.nlm.nih.gov/21496745/)]
23. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: a pilot randomised controlled trial. *PLoS One* 2018;13(12):e0209482 [FREE Full text] [doi: [10.1371/journal.pone.0209482](https://doi.org/10.1371/journal.pone.0209482)] [Medline: [30596696](https://pubmed.ncbi.nlm.nih.gov/30596696/)]
24. Plotnikoff RC, Wilczynska M, Cohen KE, Smith JJ, Lubans DR. Integrating smartphone technology, social support and the outdoor physical environment to improve fitness among adults at risk of, or diagnosed with, Type 2 diabetes: findings from the 'eCoFit' randomized controlled trial. *Prev Med* 2017 Dec;105:404-411. [doi: [10.1016/j.ypmed.2017.08.027](https://doi.org/10.1016/j.ypmed.2017.08.027)] [Medline: [28887192](https://pubmed.ncbi.nlm.nih.gov/28887192/)]
25. Spring B, Pellegrini C, McFadden HG, Pfammatter AF, Stump TK, Siddique J, et al. Multicomponent mHealth intervention for large, sustained change in multiple diet and activity risk behaviors: the make better choices 2 randomized controlled trial. *J Med Internet Res* 2018 Jun 19;20(6):e10528 [FREE Full text] [doi: [10.2196/10528](https://doi.org/10.2196/10528)] [Medline: [29921561](https://pubmed.ncbi.nlm.nih.gov/29921561/)]
26. Geiker NRW, Astrup A, Hjorth MF, Sjödin A, Pijls L, Markus CR. Does stress influence sleep patterns, food intake, weight gain, abdominal obesity and weight loss interventions and vice versa? *Obes Rev* 2018 Jan;19(1):81-97. [doi: [10.1111/obr.12603](https://doi.org/10.1111/obr.12603)] [Medline: [28849612](https://pubmed.ncbi.nlm.nih.gov/28849612/)]
27. Rutters F, Nieuwenhuizen AG, Lemmens SGT, Born JM, Westerterp-Plantenga MS. Acute stress-related changes in eating in the absence of hunger. *Obesity (Silver Spring)* 2009 Jan;17(1):72-77 [FREE Full text] [doi: [10.1038/oby.2008.493](https://doi.org/10.1038/oby.2008.493)] [Medline: [18997672](https://pubmed.ncbi.nlm.nih.gov/18997672/)]
28. Warnecke RB, Johnson TP, Chávez N, Sudman S, O'Rourke DP, Lacey L, et al. Improving question wording in surveys of culturally diverse populations. *Ann Epidemiol* 1997 Jul;7(5):334-342. [doi: [10.1016/s1047-2797\(97\)00030-6](https://doi.org/10.1016/s1047-2797(97)00030-6)] [Medline: [9250628](https://pubmed.ncbi.nlm.nih.gov/9250628/)]
29. Ekblom-Bak E, Olsson G, Ekblom ?, Ekblom B, Bergström G, Börjesson M. The daily movement pattern and fulfilment of physical activity recommendations in Swedish middle-aged adults: the SCAPIS pilot study. *PLoS One* 2015 May 13;10(5):e0126336 [FREE Full text] [doi: [10.1371/journal.pone.0126336](https://doi.org/10.1371/journal.pone.0126336)] [Medline: [25970580](https://pubmed.ncbi.nlm.nih.gov/25970580/)]
30. van der Mark M, Jonasson J, Svensson M, Linné Y, Rossner S, Lagerros YT. Older members perform better in an internet-based behavioral weight loss program compared to younger members. *Obes Facts* 2009;2(2):74-79. [doi: [10.1159/000209383](https://doi.org/10.1159/000209383)] [Medline: [20054209](https://pubmed.ncbi.nlm.nih.gov/20054209/)]
31. Ek A, Alexandrou C, Söderström E, Bergman P, Delisle Nyström C, Direito A, et al. Effectiveness of a 3-month mobile phone-based behavior change program on active transportation and physical activity in adults: randomized controlled trial. *JMIR Mhealth Uhealth* 2020 Jun 08;8(6):e18531 [FREE Full text] [doi: [10.2196/18531](https://doi.org/10.2196/18531)] [Medline: [32510462](https://pubmed.ncbi.nlm.nih.gov/32510462/)]
32. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. *J Occup Health Psychol* 2018 May 03;24(1):127-138. [doi: [10.1037/ocp0000118](https://doi.org/10.1037/ocp0000118)] [Medline: [29723001](https://pubmed.ncbi.nlm.nih.gov/29723001/)]
33. Bonn SE, Alexandrou C, Hjörleifsdóttir SK, Wiklander K, Östenson C, Löf M, et al. App-technology to increase physical activity among patients with diabetes type 2: the DiaCert-study, a randomized controlled trial. *BMC Public Health* 2018 Dec 10;18(1):119 [FREE Full text] [doi: [10.1186/s12889-018-5026-4](https://doi.org/10.1186/s12889-018-5026-4)] [Medline: [29316905](https://pubmed.ncbi.nlm.nih.gov/29316905/)]
34. Henriksson P, Sandborg J, Blomberg M, Alexandrou C, Maddison R, Silfvernel K, et al. A smartphone app to promote healthy weight gain, diet, and physical activity during pregnancy (HealthyMoms): protocol for a randomized controlled trial. *JMIR Res Protoc* 2019 Mar 01;8(3):e13011 [FREE Full text] [doi: [10.2196/13011](https://doi.org/10.2196/13011)] [Medline: [30821695](https://pubmed.ncbi.nlm.nih.gov/30821695/)]
35. Bonn SE, Hult M, Spetz K, Löf M, Andersson E, Wiren M, et al. App technology to support physical activity and intake of vitamins and minerals after bariatric surgery (the PromMera Study): protocol of a randomized controlled clinical trial. *JMIR Res Protoc* 2020 Aug 14;9(8):e19624 [FREE Full text] [doi: [10.2196/19624](https://doi.org/10.2196/19624)] [Medline: [32795990](https://pubmed.ncbi.nlm.nih.gov/32795990/)]

36. Yannitsos D, Murphy RA, Pollock P, Di Sebastiano KM. Facilitators and barriers to participation in lifestyle modification for men with prostate cancer: a scoping review. *Eur J Cancer Care (Engl)* 2020 Jan;29(1):e13193. [doi: [10.1111/ecc.13193](https://doi.org/10.1111/ecc.13193)] [Medline: [31797478](https://pubmed.ncbi.nlm.nih.gov/31797478/)]
37. Aschbrenner KA, Naslund JA, Bartels SJ. A mixed methods study of peer-to-peer support in a group-based lifestyle intervention for adults with serious mental illness. *Psychiatr Rehabil J* 2016 Dec;39(4):328-334 [FREE Full text] [doi: [10.1037/prj0000219](https://doi.org/10.1037/prj0000219)] [Medline: [27560454](https://pubmed.ncbi.nlm.nih.gov/27560454/)]
38. The Swedes and Internet 2019 (Svenskarna och internet 2019). URL: <https://svenskarnaochinternet.se/rapporter/svenskarna-och-internet-2019/> [accessed 2020-09-08]

Abbreviations

EIT: European Institute of Innovation and Technology

FINDRISC: Finnish Diabetes Risk Score

HbA_{1c}: glycated hemoglobin A1c

mHealth: mobile health

WHO: World Health Organization

Edited by A Mavragani; submitted 04.10.20; peer-reviewed by X Guo; comments to author 18.11.20; revised version received 10.01.21; accepted 01.02.22; published 21.03.22.

Please cite as:

Bonn S, Licitra G, Bellocco R, Trolle Lagerros Y

Clinical Outcomes Among Working Adults Using the Health Integrator Smartphone App: Analyses of Prespecified Secondary Outcomes in a Randomized Controlled Trial

J Med Internet Res 2022;24(3):e24725

URL: <https://www.jmir.org/2022/3/e24725>

doi: [10.2196/24725](https://doi.org/10.2196/24725)

PMID: [35311677](https://pubmed.ncbi.nlm.nih.gov/35311677/)

©Stephanie Bonn, Gabriella Licitra, Rino Bellocco, Ylva Trolle Lagerros. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 21.03.2022. 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.

Original Paper

Social Media–Delivered Patient Education to Enhance Self-management and Attitudes of Patients with Type 2 Diabetes During the COVID-19 Pandemic: Randomized Controlled Trial

Cheng Man Leong¹, MS; Ting-I Lee^{2,3}, MD, PhD; Yu-Mei Chien², MD; Li-Na Kuo^{1,4}, MS; Yu-Feng Kuo¹, MS; Hsiang-Yin Chen^{1,4}, MS, PharmD

¹Department of Clinical Pharmacy, School of Pharmacy, Taipei Medical University, Taipei, Taiwan

²Division of Endocrinology and Metabolism, Department of Internal Medicine, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

³Division of Endocrinology and Metabolism, Department of Internal Medicine, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

⁴Department of Pharmacy, Wan Fang Hospital, Taipei Medical University, Taipei, Taiwan

Corresponding Author:

Hsiang-Yin Chen, MS, PharmD

Department of Clinical Pharmacy

School of Pharmacy

Taipei Medical University

R714, 7th Floor, Health and Science Building

No.250, Wuxing St., Xinyi Dist.

Taipei, 110

Taiwan

Phone: 886 2 2736 1661 ext 6175

Fax: 886 2 2736 1661

Email: shawn@tmu.edu.tw

Abstract

Background: The use of mobile health technologies has been necessary to deliver patient education to patients with diabetes during the COVID-19 pandemic.

Objective: This open-label randomized controlled trial evaluated the effects of a diabetes educational platform—Taipei Medical University–LINE Oriented Video Education—delivered through a social media app.

Methods: Patients with type 2 diabetes were recruited from a clinic through physician referral. The social media–based program included 51 videos: 10 about understanding diabetes, 10 about daily care, 6 about nutrition care, 21 about diabetes drugs, and 4 containing quizzes. The intervention group received two or three videos every week and care messages every 2 weeks through the social media platform for 3 months, in addition to usual care. The control group only received usual care. Outcomes were measured at clinical visits through self-reported face-to-face questionnaires at baseline and at 3 months after the intervention, including the Simplified Diabetes Knowledge Scale (true/false version), the Diabetes Care Profile–Attitudes Toward Diabetes Scales, the Summary of Diabetes Self-Care Activities, and glycated hemoglobin (HbA_{1c}) levels. Health literacy was measured at baseline using the Newest Vital Sign tool. Differences in HbA_{1c} levels and questionnaire scores before and after the intervention were compared between groups. The associations of knowledge, attitudes, and self-care activities with health literacy were assessed.

Results: Patients with type 2 diabetes completed the 3-month study, with 91 out of 181 (50.3%) patients in the intervention group and 90 (49.7%) in the control group. The change in HbA_{1c} did not significantly differ between groups (intervention group: mean 6.9%, SD 0.8% to mean 7.0%, SD 0.9%, $P=.34$; control group: mean 6.7%, SD 0.6% to mean 6.7%, SD 0.7%, $P=.91$). Both groups showed increased mean knowledge scores at 12 weeks, increasing from 68.3% (SD 16.4%) to 76.7% (SD 11.7%; $P<.001$) in the intervention group and from 64.8% (SD 18.2%) to 73.2% (SD 12.6%; $P<.001$) in the control group. Positive improvements in attitudes and self-care activities were only observed in the intervention group (attitudes: mean difference 0.2, SD 0.5, $P=.001$; self-care activities: mean difference 0.3, SD 1.2, $P=.03$). A 100% utility rate was achieved for 8 out of 21 (38%) medication-related videos. Low health literacy was a significant risk factor for baseline knowledge scores in the intervention group, with an odds ratio of 2.80 (95% CI 1.28–6.12; $P=.01$); this became insignificant after 3 months.

Conclusions: The social media–based program was effective at enhancing the knowledge, attitudes, and self-care activities of patients with diabetes. This intervention was also helpful for patients with low health literacy in diabetes knowledge. The program represents a potentially useful tool for delivering diabetes education to patients through social media, especially during the COVID-19 pandemic.

Trial Registration: ClinicalTrials.gov NCT04876274; <https://clinicaltrials.gov/ct2/show/results/NCT04876274>

(*J Med Internet Res* 2022;24(3):e31449) doi:[10.2196/31449](https://doi.org/10.2196/31449)

KEYWORDS

diabetes; COVID-19; education; video; social media; health literacy; self-care activity; type 2 diabetes; attitude; mHealth

Introduction

Background

The development of effective diabetes education programs for patient self-management has been a challenge for health care professionals. The sudden outbreak of COVID-19 leading to the pandemic has created further difficulty by limiting face-to-face diabetes education. The complicated pathology of diabetes requires not only pharmacotherapy, but also firm patient engagement with daily self-care. Studies have demonstrated that patients with better disease-related knowledge, attitudes, practice, and self-efficacy have better glycemic control [1]. However, it is extremely difficult for patients with diabetes to maintain a healthy lifestyle and effective self-management without professional assistance. The excessive workload in health care settings during the COVID-19 pandemic has limited the time available to provide sufficient patient education. The increasing prevalence rates of diabetes worldwide reflects the unmet need for health education, which calls for innovative and effective mobile health (mHealth) educational programs.

The content of diabetes education is very complicated, as demonstrated by Diabetes Self-Management Education and Support services [2] and the American Association of Diabetes Educators 7 Self-Care Behaviors (AADE7) [3]. To cover all aspects required for day-to-day living with diabetes, it is recommended that education include healthy eating, regular physical activities, self-monitoring of blood glucose, compliance with medications, problem-solving skills, healthy coping skills, and other risk-reducing behaviors [3]. These aspects have been theoretically proven to effectively enhance health outcomes [4-8]. However, educating patients on all of these topics might not be practical in busy clinical settings. The time needed to teach diabetes self-care is reported to be around 4 hours per patient, which has been difficult to achieve in clinical practice during the COVID-19 pandemic [9]. Remote learning using advanced technologies may potentially provide a key solution to address the pressures created by the burden of face-to-face education.

mHealth technology has increasingly been integrated into health care to meet the demands of diabetic care during the COVID-19 pandemic. It has been shown to be effective at enhancing health outcomes, such as medication adherence, glycated hemoglobin (HbA_{1c}) levels, and self-management [10-18]. Studies have shown that multimedia education using videos is more effective than written information in terms of engagement and information uptake [19,20]. Videos can deliver information through visual

and audio elements and require less cognitive effort for information processing [19,20]. However, no previous study has attempted to develop a video program for diabetes education based on social media with a two-way communication component.

Objectives

In this study, we developed an educational program based on the AADE7 [3]: Taipei Medical University–LINE Oriented Video Education (TMU-LOVE). The social media platform LINE (LINE Corporation) was employed in this study. The platform is one of the most popular and user-friendly social media platforms in Taiwan, with a high acceptance rate of up to 88% among the population between the ages of 16 and 64 years [21]. This randomized controlled trial (RCT) was conducted during the COVID-19 pandemic. The first aim was to evaluate the effectiveness of the social media–based education program in regard to changes in patients' HbA_{1c} levels, knowledge, attitudes, and self-care activities before and after the intervention. Another aim was to study the effects of this social media–based education program on changes in knowledge by diabetes patients with high and low levels of health literacy.

Methods

Study Design

This study was an open-label RCT and was conducted between July 2020 and January 2021. The study was registered at ClinicalTrials.gov (NCT04876274).

Ethics Approval

This study was approved by the Taipei Medical University (TMU)–Joint Institutional Review Board (No. N201905088). The study followed the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) guidelines [22].

Participants

Participants were recruited by physicians at the Endocrinology and Metabolism Clinic of Wan-Fang Hospital, Taipei, Taiwan. Invitations to join the study were extended to all patients with diabetes, aged 20 years or older, who had at least one HbA_{1c} measurement that was equal to or greater than 6% in the past 6 months and possessed a smartphone with the LINE app installed. Patients who had gestational diabetes, had cognitive impairment, or were controlling their diabetes through dietary control without medication were excluded. Based on a power of 80% and a 2-sided α level of .05, the required sample size was estimated

to be at least 128 participants using G*Power (version 3.1; Heinrich-Heine-Universität Düsseldorf) [23]. To adjust for an estimated nonresponse rate of 20%, we aimed for a final sample size of at least 160 participants, with 80 patients in each group.

Eligible participants were asked to complete a written consent form before joining the study. Demographic information (ie, age, sex, educational level, income, medications, and personal history) and health status (ie, diagnosis time and BMI) were obtained using electronic medical records. Pharmacists allocated participants to either the control group or the intervention group at a 1:1 ratio according to a random sequence, which was generated by a digital program before launching the study. The patients in both groups were scheduled to return to the clinic at the end of the 3-month study period, where their chronic prescription medications were re-evaluated. Diabetes health education for the control group comprised provision of the usual diabetes health care, including patient consultations with physicians, access to nurses in the outpatient services, as well as medication consultations with pharmacists upon receiving prescriptions. Physicians also referred patients for consultations with certified diabetes educators (CDEs) when necessary.

Intervention

The patients allocated to the intervention group were granted access to the TMU-LOVE program through a QR code (Figure S1 in [Multimedia Appendix 1](#)), in addition to receiving the usual health care. This program was developed by the TMU School of Pharmacy and is illustrated in [Figure 1](#). In the first phase of development, a panel consisting of an endocrinologist, two pharmacists, and a pharmacy professor developed an outline of the program according to the guidelines of the Taiwanese Association of Diabetes Educators and seven key points in the AADE7 [3]. The outline contained five categories: understanding diabetes, daily care, nutrition care, diabetes drugs, and quizzes. The elements of each section were also determined.

A group of pharmacy students joined the second phase to create the videos. The students started to write scenarios and discussed each video with the pharmacy professor and pharmacists. After checking the evidence for the scenarios, the students produced

animated or filmed videos. The videos were reviewed and revised by the panel until they met the learning goals in the outline. A total of 51 videos were produced: 10 videos were about understanding diabetes, 10 were about daily care, 6 were about nutrition care, 21 were about diabetes drugs, and 4 were about diabetes knowledge-related quizzes (Figure S2 in [Multimedia Appendix 1](#)). The medications described in the program were based on the formulary at Wan-Fang Hospital, including metformin, acarbose, dipeptidyl peptidase-IV inhibitors, meglitinide, sulfonylurea, thiazolidinedione, sodium-glucose cotransporter-2 inhibitors, and insulin. All videos are listed in [Multimedia Appendix 2](#) and are available at the TMU School of Pharmacy's website [24].

The menu for the educational platform was composed of six icons; these represented the five video categories and a FAQ (frequently asked questions) section (Figure S3 in [Multimedia Appendix 1](#)). The program was also equipped with a one-on-one chat room, graphic messages, and the ability to direct patients to certain websites [25]. The study employed the messaging feature of the LINE social media platform to send educational videos and communicate with patients by text message, voice, or video call. Patients could also click on the six icons in the menu to connect to the website of the program. The patients could use LINE's messaging feature to ask questions, and pharmacists could answer them through text messages or voice calls via the social media platform.

The program was designed to last for a study duration of 3 months (ie, 12 weeks), with two or three videos sent every week and a care message sent every 2 weeks to the patients in the intervention group. All patients received the same videos except for the videos on medications. Videos regarding the basic understanding of diabetes were delivered in weeks 1 to 4. Videos on daily care were first delivered in week 5, and six nutrition care videos were scheduled over six different weeks. Each patient received the videos for their individual medications in weeks 4, 6, 7, 11, and 12. Four quizzes were scheduled to be delivered in weeks 3, 6, 9, and 12 ([Figure 2](#)). Patients could also access all of the videos through the social media platform whenever they desired.

Figure 1. Program development. The number of videos in each category are reported within parentheses. AADE7: American Association of Diabetes Educators 7 Self-Care Behaviors; FAQ: frequently asked questions; TADE: Taiwanese Association of Diabetes Educators; TMU-LOVE: Taipei Medical University–LINE Oriented Video Education.

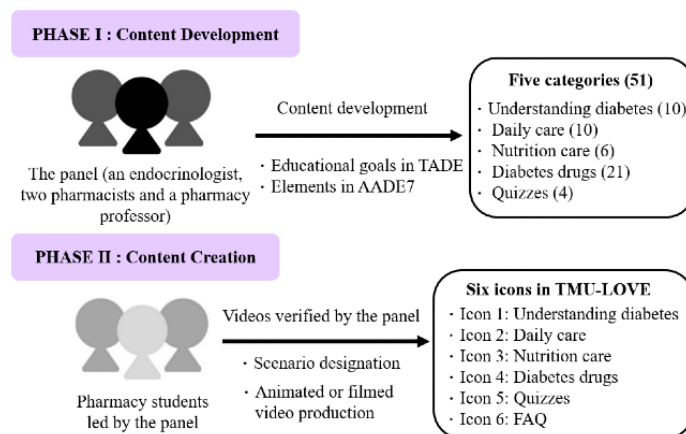
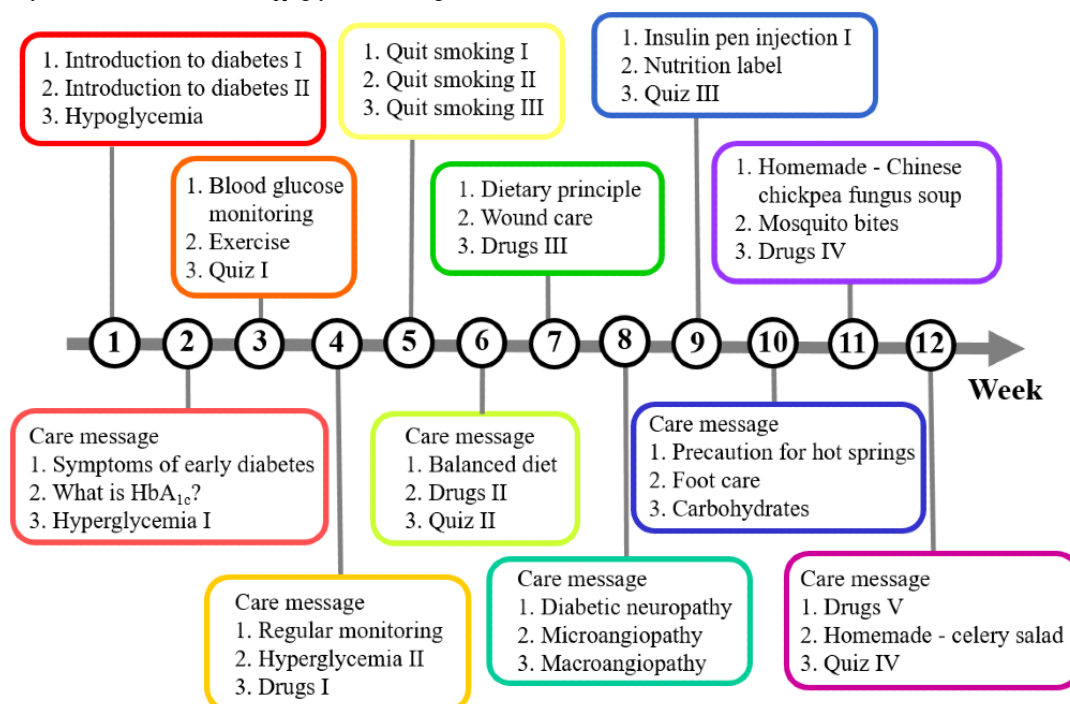


Figure 2. Delivery schedule of videos. HbA_{1c}: glycated hemoglobin.

Outcomes and Measures

Measures for Blood Sugar Control

The outcome for blood sugar control was the change of HbA_{1c} level from baseline to month 3. HbA_{1c} level results were obtained from the electronic medical records for all potential participants before the allocation and determination of eligibility. The participants were requested to take an HbA_{1c} test at 3 months, and all HbA_{1c} level results during the 3-month study period were examined for the analysis.

Knowledge, Attitudes, and Self-care Activities Toward Diabetes

The Chinese version of the Simplified Diabetes Knowledge Scale (SDKS; true/false version) [26] was used to measure diabetes patients' knowledge. This version of the scale contains 24 items and was validated in previous studies [26,27].

Patients' attitudes toward diabetes were assessed using the Chinese version of the Diabetes Care Profile—Attitudes Toward Diabetes Scales (DCP-ATDS) [26]. This questionnaire was previously validated and translated [26,28]. It has 17 items that are divided into the following sections: positive attitudes, negative attitudes, self-care ability, self-care adherence, and importance of care. Each item is graded on a 5-point Likert scale using the following response options: "strongly agree," "agree," "neutral," "disagree," or "strongly disagree."

Self-care activities were measured using the Chinese version of the Summary of Diabetes Self-Care Activities (SDSCA) [29], which consists of 10 items and has been validated [29,30]. In the SDSCA, patients were asked how many days per week they had performed the correct behaviors regarding medication adherence, diet, exercise, self-monitoring of blood glucose, and foot care.

Health Literacy

To determine whether the social media-based program could be applied to patients with different levels of health literacy, the effects of health literacy on SDKS, SDSCA, and DCP-ATDS outcomes were compared in the intervention and control groups. We assessed health literacy at baseline using the Newest Vital Sign (NVS) tool, which was developed and translated in previous studies [31,32]. This scale consists of six questions based on the nutrition label of an ice cream product in order to evaluate both the reading and numeracy of the patients. The patients were categorized as having high or low health literacy based on a cutoff point of 2 for the sum of the NVS score.

Utility Rate of Educational Videos

The utility rate was used to represent the frequency of watching videos in each of the five categories. The utility rate for accessing educational videos was measured using the number of views of each video divided by the number of target participants. Utility rates of videos with a calculated rate above 100% were counted as 100%.

Statistical Analysis

Statistical analyses were conducted using SPSS Statistics for Windows (version 28.0; IBM Corp). The study was performed using intention-to-treat analysis. Multiple imputation was employed to manage missing values [33-35]. All tests were 2-tailed, with a significance level of .05. Differences in baseline characteristics were examined using descriptive analysis. The Kolmogorov-Smirnov test was used to verify whether the variables had normal distributions. According to the results of the Kolmogorov-Smirnov test, a Wilcoxon signed-rank test or paired *t* tests were used to examine the differences in the pretest and posttest scores within a group, while a Mann-Whitney *U* test or unpaired *t* tests were performed to compare the differences between groups. An ordinal logistic regression model

was applied to evaluate the association of health literacy with knowledge, attitudes, and self-care activities.

Results

Participant Characteristics

A total of 246 patients from the Endocrinology and Metabolism Clinic of Wan-Fang Hospital were screened for this study. Patients who were excluded from the study included 10 patients with HbA_{1c} levels less than 6%, 3 patients who did not complete the informed consent form, and 52 patients who declined to participate. In total, 181 patients were randomized equally into

the control group (n=90, 49.7%) or the intervention group (n=91, 50.3%), as shown in Figure 3.

Table 1 shows the demographic characteristics of the patients, which indicate that there were no significant differences observed between groups. The percentage of patients receiving CDE care in the two groups differed insignificantly. The mean age of the participants was 58.6 (SD 11.6) years, and there were 57 males out of 90 participants (63.3%) in the control group and 67 males out of 91 participants (73.6%) in the intervention group. More than half of the participants were highly educated. The mean HbA_{1c} level was 6.8% (SD 0.7%), and the mean health literacy level was 1.7 (SD 1.9).

Figure 3. The CONSORT flow diagram. CONSORT: Consolidated Standards of Reporting Trials; TMU-LOVE: Taipei Medical University–LINE Oriented Video Education.

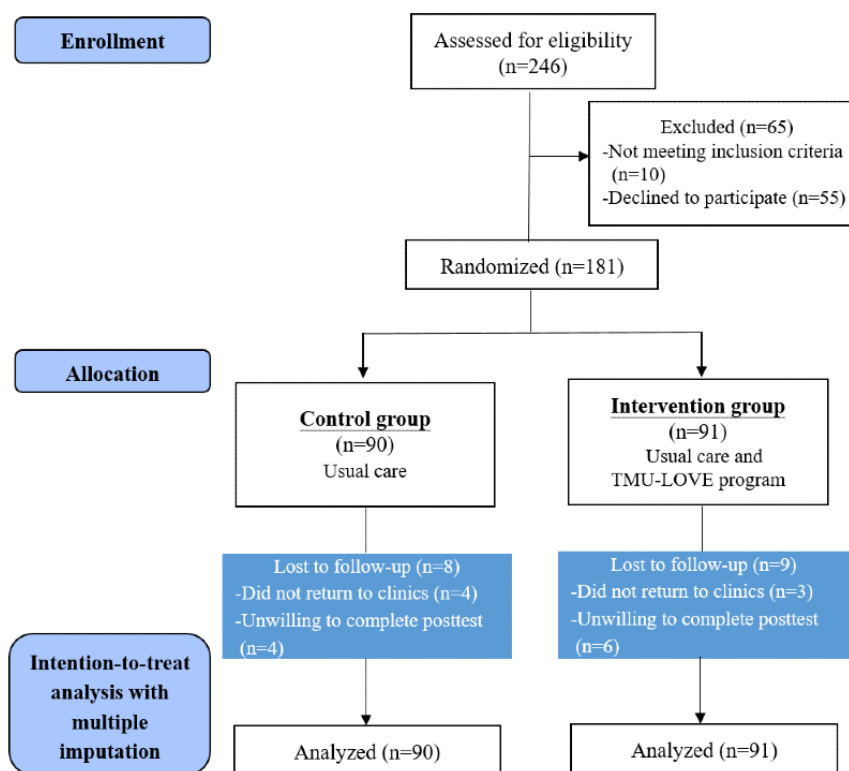


Table 1. Baseline characteristics of the participants.

Variable	All participants (N=181)	Control group (n=90)	Intervention group (n=91)	P value
Age (years), mean (SD)	58.6 (11.6)	58.1 (11.9)	59.0 (11.4)	.60
Male, n (%)	124 (68.5)	57 (63.3)	67 (73.6)	.15
BMI ^a , mean (SD)	26.4 (4.6)	26.7 (5.0)	26.1 (4.2)	.39
Educational level, n (%)				
Junior high school or below	27 (14.9)	13 (14.4)	14 (15.4)	.93 ^b
Senior high school	50 (27.6)	24 (26.7)	26 (28.6)	
College or higher	104 (57.5)	53 (58.9)	51 (56.0)	
Unemployed, n (%)	76 (42.0)	40 (44.4)	36 (39.6)	.55
Living alone, n (%)	6 (3.3)	3 (3.3)	3 (3.3)	.60
Annual income (NT\$^c), n (%)				
Unwaged	58 (32.0)	30 (33.3)	28 (30.8)	.48
<500,000	43 (23.8)	25 (27.8)	18 (19.8)	
500,000-1,000,000	48 (26.5)	21 (23.3)	27 (29.7)	
>1,000,000	32 (17.7)	14 (15.6)	18 (19.8)	
Medication type, n (%)				
Oral medication only	140 (77.3)	72 (80)	68 (74.7)	.06
Insulin only	9 (5.0)	1 (1.1)	8 (8.8)	
Oral medication and insulin	32 (17.7)	17 (18.9)	15 (16.5)	
Diagnosis time (years), mean (SD)	10.3 (8.5)	10.3 (8.1)	10.4 (9.0)	.91
HbA _{1c} ^d level (%), mean (SD)	6.8 (0.7)	6.7 (0.6)	7.0 (0.8)	.07
Smoking, n (%)	25 (13.8)	15 (16.7)	10 (11.0)	.23
Alcohol consumption, n (%)	65 (35.9)	29 (32.2)	36 (39.6)	.78
SMBG ^e , n (%)	79 (43.6)	42 (46.7)	37 (40.7)	.46
Physical exercise, n (%)	133 (73.5)	63 (70)	70 (76.9)	.77
Receiving CDE ^f education within 3 months, n (%)	43 (23.8)	20 (22.2)	23 (25.3)	.73
NVS ^g questionnaire total score, mean (SD)	1.7 (1.9)	1.8 (2.0)	1.7 (1.8)	.71

^aBMI is calculated as weight in kilograms divided by height in meters squared.

^bP values for a group are reported in the top row of that group.

^c\$1 NT = \$0.035 USD.

^dHbA_{1c}: glycated hemoglobin.

^eSMBG: self-monitoring blood glucose.

^fCDE: certified diabetes educator.

^gNVS: Newest Vital Sign; questionnaire total scores range from 0 to 6, with a higher score indicating a favorable state.

Outcomes

Measures for Blood Sugar Control

After 3 months of the intervention, the mean HbA_{1c} level changed insignificantly from 6.9% (SD 0.8%) to 7% (SD 0.9%;

$P=.34$) in the intervention group and from 6.7% (SD 0.6%) to 6.7% (SD 0.7%; $P=.91$) in the control group (Table 2).

Table 2. Clinical and nonclinical outcomes after the intervention.

Variable	Within groups ^a			Between groups ^b			
	Baseline, mean (SD)	3 months, mean (SD)	<i>P</i> value	Mean difference (SD)	Mean difference (SD)	95% CI	<i>P</i> value
HbA_{1c}^c (%)							
Intervention group (n=91)	6.9 (0.8)	7.0 (0.9)	.34	0.07 (0.7)	0.06 (0.1) ^d	-0.25 to 0.14 ^d	.57 ^d
Control group (n=90)	6.7 (0.6)	6.7 (0.7)	.91	0.01 (0.7)			
SDKS^e (% correctness)							
Intervention group (n=91)	68.3 (16.4)	76.7 (11.7)	<.001	8.4 (14.7)	0.01 (2.3)	-4.52 to 4.53	.99
Control group (n=90)	64.8 (18.2)	73.2 (12.6)	<.001	8.4 (16.1)			
DCP-ATDS^f							
Intervention group (n=91)	3.6 (0.4)	3.8 (0.5)	.001	0.2 (0.5)	0.2 (0.06)	-0.32 to -0.07	.003
Control group (n=90)	3.7 (0.4)	3.7 (0.5)	.58	-0.02 (0.4)			
SDSCA^g							
Intervention group (n=91)	3.7 (1.3)	4.0 (1.2)	.03	0.3 (1.2)	0.3 (0.2)	-0.65 to -0.02	.04
Control group (n=90)	3.9 (1.4)	3.9 (1.5)	.52	-0.07 (1.2)			

^aPaired *t* tests (2-tailed) were performed for HbA_{1c} levels, DCP-ATDS scores, and SDSCA scores; a Wilcoxon signed-rank test was performed for percentage of correct SDKS questions.

^bUnpaired *t* tests (2-tailed) were performed for HbA_{1c} levels, DCP-ATDS scores, and SDSCA scores; a Mann-Whitney *U* test was performed for percentage of correct SDKS questions.

^cHbA_{1c}: glycated hemoglobin.

^dValues for a group are reported in the top row of that group.

^eSDKS: Simplified Diabetes Knowledge Scale (true/false version). Percent correctness ranges from 0% to 100%, and a higher score indicates a favorable state.

^fDCP-ATDS: Diabetes Care Profile—Attitudes Toward Diabetes Scales. Mean scores range from 1 to 5, and a higher score indicates a favorable state; pretest scores were normally distributed, based on the Kolmogorov-Smirnov test.

^gSDSCA: Summary of Diabetes Self-Care Activities. Mean scores range from 0 to 7, and a higher score indicates a favorable state; pretest scores were normally distributed, based on the Kolmogorov-Smirnov test.

Knowledge, Attitudes, and Self-care Activities Toward Diabetes

Both groups experienced significant improvements in SDKS mean scores, from 68.3% (SD 16.4%) to 76.7% (SD 11.7%; *P*<.001) in the intervention group and from 64.8% (SD 18.2%) to 73.2% (SD 12.6%; *P*<.001) in the control group. However, the difference in knowledge scores between the two groups was not significant (intervention, 8.4% vs control, 8.4%; *P*=.85). The scores for each item of the SDKS are shown in [Multimedia Appendix 3](#).

The intervention group exhibited significant growth in the overall attitude mean score, from 3.6 (SD 0.4) to 3.8 (SD 0.5; *P*=.001), whereas no significant change was observed in the control group, from 3.7 (SD 0.4) to 3.7 (SD 0.5; *P*=.58). The difference in the change of DCP-ATDS scores between baseline and follow-up was significant between groups (mean difference 0.2, 95% CI -0.32 to -0.07; *P*=.003). Detailed data for the DCP-ATDS are presented in [Multimedia Appendix 4](#).

Regarding self-care activities measured by the SDSCA, significant improvement was observed in the intervention group

only, from a mean score of 3.7 (SD 1.3) to 4.0 (SD 1.2; *P*=.03); the improvement in scores was significantly higher in comparison to the control group (mean 0.3, SD 0.2; *P*=.04). Tables S1 and S2 in [Multimedia Appendix 5](#) show the SDSCA scores and the top five reasons mentioned by patients for finding it difficult to engage in self-care activities. In further analyses, patients with and without improvement in SDSCA scores after 3 months had no significant differences in any of DCP-ATDS subscales, including the subscales of positive attitudes and negative attitudes ([Multimedia Appendix 6](#)).

Impact of Health Literacy

[Table 3](#) shows the effects of health literacy on knowledge, attitudes, and self-care activities associated with the SDKS. At baseline, the odd ratios (ORs) of health literacy were 2.8 (95% CI 1.28-6.12; *P*=.01) for the intervention group and 5.43 (95% CI 2.41-12.25; *P*<.001) for the control group. The social media-based education intervention eased the influence of health literacy on SDKS results, which disappeared in the intervention group after 3 months, with a *P* value of .75. The OR remained significant in the control group after 3 months (OR 2.43, 95% CI 1.14-5.19; *P*=.02).

Table 3. Effects of health literacy on knowledge, attitudes, and self-care activities.

Variables	Intervention group (n=91)				Control group (n=90)			
	Baseline		3 months		Baseline		3 months	
	OR ^a (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
SDKS ^b	2.80 (1.28-6.12)	.01	0.88 (0.41-1.91)	.75	5.43 (2.41-12.25)	<.001	2.43 (1.14-5.19)	.02
DCP-ATDS ^c	1.60 (0.75-3.41)	.23	1.45 (0.66-3.21)	.36	0.77 (0.37-1.62)	.49	0.90 (0.42-1.91)	.64
SDSCA ^d	0.44 (0.21-0.95)	.04	0.76 (0.36-1.62)	.48	0.95 (0.45-2.00)	.89	0.92 (0.43-1.95)	.82

^aOR: odds ratio.

^bSDKS: Simplified Diabetes Knowledge Scale (true/false version).

^cDCP-ATDS: Diabetes Care Profile–Attitudes Toward Diabetes Scales.

^dSDSCA: Summary of Diabetes Self-Care Activities.

Utility Rate of Educational Videos

The utility rates for drug-related videos were the highest, with 8 out of 21 (38%) videos achieving a 100% utility rate. For the category of understanding diabetes, only 4 out of 10 (40%) videos had a utility rate above 50%. Among the 10 videos

related to daily care, only 2 (20%) had a utility rate above 50%. Only 1 nutrition care video out of 6 (17%) and 1 quiz video out of 4 (25%) was highly used (Figure 4). Interestingly, the utility rate of the first quiz reached 69.2%, but the utility rates were significantly lower for the following three quizzes, ranging from 27.5% to 30.8%.

Figure 4. Utility rate for each video in each of the five categories. Rates are shown on each bar. HbA_{1c}: glycated hemoglobin.



Discussion

Principal Findings

This study demonstrated that the delivery of a social media-based education program during the COVID-19 pandemic improved the knowledge, attitudes, and self-care activities of patients with type 2 diabetes. Among the 51 videos in the social media intervention, the utility rates of the drug-related videos were the highest, which implies a demand for medication information among patients. Interestingly, the results also revealed that the program overcame the negative effects of low health literacy, as it was no longer a significant factor in knowledge improvement after the intervention. However, the significant influence of low health literacy remained in the control group after 3 months. The program demonstrated the usefulness of social media-based diabetes education, even among those who have limited health literacy.

The attitude and practice improvement observed in this study, which was not always present with knowledge enhancement, requires stronger motivation and patient engagement. The videos in this study covered all elements of the AADE7 and drug information, resulting in reasonable knowledge improvement. Previous mHealth studies showed conflicting results regarding attitudes. One study showed that the use of mobile text messaging did not significantly benefit diabetes patients' attitudes [36], whereas another study showed that the use of mobile texts, voice messages, and animations improved patient attitudes [37]. All videos in this study were presented with a mindset of fostering growth and positive attitudes to help patients fight against diabetes. The positive encouragement was fundamentally important for changing the mindset of patients, as the chronic nature of disease commonly decreases patients' perseverance.

Changes in self-care activities were even more difficult to achieve, requiring a more individualized, patient-centered approach. Patient interviews in this study revealed that a group characterized by having a high workload experienced difficulties in improving their self-care activities. They ate high-fat meals and engaged in little to no exercise. Interestingly, this group had the same attitude level as patients who showed improvements in self-care activities. This finding was consistent with previous studies [38,39], indicating that many patients with diabetes have good attitudes but poor practice. Four videos with easy-to-make recipes in this study might have motivated patients to take the initiative toward self-care. Individualized discussion is essential for assisting patients with high workloads to overcome difficulties in changing their lifestyle [40].

This study identified a key lesson for operating a successful mHealth education program for patients with diabetes during the COVID-19 pandemic. Worldwide health care and patient education shifted quickly from in-person to remote online approaches due to physical distancing and social isolation restrictions [41-43]. Online telemedicine outpatient services have become a popular trend in health care settings [44,45]. Revision of the content of the social media program by adding the facility to have more frequent online discussions could be

one solution to optimize diabetes health education during the COVID-19 pandemic.

This study demonstrated that patient-oriented language and animated videos were effective at being accepted by patients with inadequate health literacy. Health literacy decreases with age due to degeneration of cognitive and intellectual functions, and is positively related to a patient's level of education [46-48]. Consistent with other studies [49,50], the level of health literacy is generally low in Taiwan. Even among our highly educated populations, the mean NVS score was still less than 2, indicating a high likelihood of limited literacy. Possible reasons might be that health education is not being emphasized in the high school curriculum or adopted as a university entrance requirement in Taiwan. The need to strengthen basic health education in formal educational settings should be considered by policy makers [50]. This study indicates that video approaches may provide an effective solution to address the generally low health literacy levels among the population and an aging society in overloaded health care settings.

Drug-related education in diabetes is an unmet clinical need. A previous study revealed that medication-related features represented four of the top 10 most demanded features by patients in an mHealth program [51]. This finding is consistent with our observation that the video utility rates were the highest for the medication category (Figure 4). According to global statistics from the Certified Diabetes Care and Education Specialist organization, pharmacists only represented 7% of all CDEs in 2020 [52]. Most CDEs in clinical settings are not pharmacists. Current mHealth education programs developed for diabetes have mainly focused on basic information, patient self-care skills, and nutrition. Development of mHealth applications with more sophisticated medication education and opportunities for patients to communicate with pharmacists could potentially help to fulfill this patient need.

Clinical Implications

The social media-based education program in this study was effective at assisting all health care professionals who provide diabetes medication education in clinical settings. When diabetes education is provided by a nonpharmacist CDE, the program could create a link to provide medication information and a channel for consultations with pharmacists. It could save pharmacists time by delivering basic medication knowledge in videos and resolving patients' general concerns. Remote delivery of videos through a social media platform provides an easy solution to help patients understand information in situations when face-to-face health education is not possible, such as during the COVID-19 pandemic. The intervention employed in this study could also be adapted to provide patient education for other chronic diseases in order to enhance telemedicine communication. Furthermore, it could be sustainably integrated into diabetes education programs run by nonpharmacist CDEs as well as pharmacists during and beyond COVID-19.

Limitations

Our study has several limitations. It was a single-center study, which may have led to population bias and, thus, restricted its generalizability. The intervention only lasted 12 weeks, so we

were unable to assess long-term outcomes. The patients' attitude, knowledge, and self-care activity scores were collected by self-report, which may have created recall bias. Moreover, we could only determine that patients clicked on the videos, but we could not confirm how long and how attentively they watched them. Finally, the economic value of the social media-based program was not evaluated. Future studies covering populations that are more diverse with a longer study duration and careful analysis of digital learning metrics are needed to validate the effects of the social media-based program among patients with type 2 diabetes.

Conclusions

This study has demonstrated the potential value of a social media-oriented video education program as a useful tool for diabetes care during the COVID-19 pandemic. It was found to have been effective in improving patients' knowledge, attitudes, and self-care activities. The video program also overcame the issue related to inadequate health literacy, which is a time-consuming challenge in traditional face-to-face education [53]. Due to the integration of medication information and two-way communication, the social media-based program represents an innovative and efficient strategy for enhancing patients' attitudes, knowledge, and self-care during the COVID-19 pandemic and in the future.

Acknowledgments

This work was supported by funding from the Taiwan Ministry of Health and Welfare (grants MOHW109-TDU-B-212-114007 and MOHW110-TDU-B-212-124007). The sponsors of the study were not involved the study design, system development, or data evaluation.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The content of the Taipei Medical University–LINE Oriented Video Education (TMU-LOVE) program.

[[DOCX File , 1141 KB - jmir_v24i3e31449_app1.docx](#)]

Multimedia Appendix 2

List of all videos.

[[DOCX File , 46011 KB - jmir_v24i3e31449_app2.docx](#)]

Multimedia Appendix 3

Correctness of the Simplified Diabetes Knowledge Scale (SDKS; true/false version).

[[DOCX File , 17 KB - jmir_v24i3e31449_app3.docx](#)]

Multimedia Appendix 4

Scores for the Diabetes Care Profile–Attitudes Toward Diabetes Scales (DCP-ATDS).

[[DOCX File , 17 KB - jmir_v24i3e31449_app4.docx](#)]

Multimedia Appendix 5

Summary of Diabetes Self-Care Activities (SDSCA) scores and reasons for finding it difficult to engage in self-care activities.

[[DOCX File , 16 KB - jmir_v24i3e31449_app5.docx](#)]

Multimedia Appendix 6

Comparison in subscale scores for the Diabetes Care Profile–Attitudes Toward Diabetes Scales (DCP-ATDS).

[[DOCX File , 15 KB - jmir_v24i3e31449_app6.docx](#)]

Multimedia Appendix 7

CONSORT-EHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 1304 KB - jmir_v24i3e31449_app7.pdf](#)]

References

1. Chvala CA, Sherr D, Lipman RD. Diabetes self-management education for adults with type 2 diabetes mellitus: A systematic review of the effect on glycemic control. *Patient Educ Couns* 2016 Jun;99(6):926-943 [[FREE Full text](#)] [doi: [10.1016/j.pec.2015.11.003](https://doi.org/10.1016/j.pec.2015.11.003)] [Medline: [26658704](https://pubmed.ncbi.nlm.nih.gov/26658704/)]

2. Beck J, Greenwood DA, Blanton L, Bollinger ST, Butcher MK, Condon JE, et al. 2017 National Standards for Diabetes Self-Management Education and Support. *Diabetes Educ* 2020 Feb 24;46(1):46-61. [doi: [10.1177/0145721719897952](https://doi.org/10.1177/0145721719897952)] [Medline: [31874594](https://pubmed.ncbi.nlm.nih.gov/31874594/)]
3. American Association of Diabetes Educators. An effective model of diabetes care and education: Revising the AADE7 self-care behaviors. *Diabetes Educ* 2020 Apr;46(2):139-160. [doi: [10.1177/0145721719894903](https://doi.org/10.1177/0145721719894903)] [Medline: [31928334](https://pubmed.ncbi.nlm.nih.gov/31928334/)]
4. Tshiananga JK, Kocher S, Weber C, Erny-Albrecht K, Berndt K, Neeser K. The effect of nurse-led diabetes self-management education on glycosylated hemoglobin and cardiovascular risk factors: A meta-analysis. *Diabetes Educ* 2012 Nov 23;38(1):108-123. [doi: [10.1177/0145721711423978](https://doi.org/10.1177/0145721711423978)] [Medline: [22116473](https://pubmed.ncbi.nlm.nih.gov/22116473/)]
5. Diabetes Control and Complications Trial Research Group. Effect of intensive diabetes treatment on the development and progression of long-term complications in adolescents with insulin-dependent diabetes mellitus: Diabetes Control and Complications Trial. *J Pediatr* 1994 Aug;125(2):177-188. [doi: [10.1016/s0022-3476\(94\)70190-3](https://doi.org/10.1016/s0022-3476(94)70190-3)]
6. He X, Li J, Wang B, Yao Q, Li L, Song R, et al. Diabetes self-management education reduces risk of all-cause mortality in type 2 diabetes patients: A systematic review and meta-analysis. *Endocrine* 2017 Mar;55(3):712-731. [doi: [10.1007/s12020-016-1168-2](https://doi.org/10.1007/s12020-016-1168-2)] [Medline: [27837440](https://pubmed.ncbi.nlm.nih.gov/27837440/)]
7. Tang TS, Funnell MM, Oh M. Lasting effects of a 2-year diabetes self-management support intervention: Outcomes at 1-year follow-up. *Prev Chronic Dis* 2012 Jun;9:E109 [FREE Full text] [doi: [10.5888/pcd9.110313](https://doi.org/10.5888/pcd9.110313)] [Medline: [22677159](https://pubmed.ncbi.nlm.nih.gov/22677159/)]
8. Cochran J, Conn VS. Meta-analysis of quality of life outcomes following diabetes self-management training. *Diabetes Educ* 2008 Sep;34(5):815-823 [FREE Full text] [doi: [10.1177/0145721708323640](https://doi.org/10.1177/0145721708323640)] [Medline: [18832286](https://pubmed.ncbi.nlm.nih.gov/18832286/)]
9. Shubrook JH, Brannan GD, Wapner A, Klein G, Schwartz FL. Time needed for diabetes self-care: Nationwide survey of certified diabetes educators. *Diabetes Spectr* 2018 Aug 06;31(3):267-271 [FREE Full text] [doi: [10.2337/ds17-0077](https://doi.org/10.2337/ds17-0077)] [Medline: [30140143](https://pubmed.ncbi.nlm.nih.gov/30140143/)]
10. Or CK, Liu K, So MKP, Cheung B, Yam LYC, Tiwari A, et al. Improving self-care in patients with coexisting type 2 diabetes and hypertension by technological surrogate nursing: Randomized controlled trial. *J Med Internet Res* 2020 Mar 27;22(3):e16769 [FREE Full text] [doi: [10.2196/16769](https://doi.org/10.2196/16769)] [Medline: [32217498](https://pubmed.ncbi.nlm.nih.gov/32217498/)]
11. Lim S, Kang SM, Shin H, Lee HJ, Won Yoon J, Yu SH, et al. Improved glycemic control without hypoglycemia in elderly diabetic patients using the ubiquitous healthcare service, a new medical information system. *Diabetes Care* 2011 Feb;34(2):308-313 [FREE Full text] [doi: [10.2337/dc10-1447](https://doi.org/10.2337/dc10-1447)] [Medline: [21270188](https://pubmed.ncbi.nlm.nih.gov/21270188/)]
12. 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](https://doi.org/10.2196/mhealth.3882)] [Medline: [25499872](https://pubmed.ncbi.nlm.nih.gov/25499872/)]
13. Abu-Saad K, Murad H, Barid R, Olmer L, Ziv A, Younis-Zeidan N, et al. Development and efficacy of an electronic, culturally adapted lifestyle counseling tool for improving diabetes-related dietary knowledge: Randomized controlled trial among ethnic minority adults with type 2 diabetes mellitus. *J Med Internet Res* 2019 Oct 16;21(10):e13674 [FREE Full text] [doi: [10.2196/13674](https://doi.org/10.2196/13674)] [Medline: [31621640](https://pubmed.ncbi.nlm.nih.gov/31621640/)]
14. Yeung DL, Alvarez KS, Quinones ME, Clark CA, Oliver GH, Alvarez CA, et al. *J Am Pharm Assoc (2003)* 2017;57(1):30-37 [FREE Full text] [doi: [10.1016/j.japh.2016.08.012](https://doi.org/10.1016/j.japh.2016.08.012)] [Medline: [27816544](https://pubmed.ncbi.nlm.nih.gov/27816544/)]
15. Vaala SE, Hood KK, Laffel L, Kumah-Crystal YA, Lybarger CK, Mulvaney SA. Use of commonly available technologies for diabetes information and self-management among adolescents with type 1 diabetes and their parents: A web-based survey study. *Interact J Med Res* 2015 Dec 29;4(4):e24 [FREE Full text] [doi: [10.2196/ijmr.4504](https://doi.org/10.2196/ijmr.4504)] [Medline: [26715191](https://pubmed.ncbi.nlm.nih.gov/26715191/)]
16. Shetty VB, Soon WHK, Roberts AG, Fried L, Roby HC, Smith GJ, et al. A novel mobile health app to educate and empower young people with type 1 diabetes to exercise safely: Prospective single-arm mixed methods pilot study. *JMIR Diabetes* 2021 Oct 14;6(4):e29739 [FREE Full text] [doi: [10.2196/29739](https://doi.org/10.2196/29739)] [Medline: [34647896](https://pubmed.ncbi.nlm.nih.gov/34647896/)]
17. Wu Y, Yao X, Vespasiani G, Nicolucci A, Dong Y, Kwong J, et al. Mobile app-based interventions to support diabetes self-management: A systematic review of randomized controlled trials to identify functions associated with glycemic efficacy. *JMIR Mhealth Uhealth* 2017 Mar 14;5(3):e35 [FREE Full text] [doi: [10.2196/mhealth.6522](https://doi.org/10.2196/mhealth.6522)] [Medline: [28292740](https://pubmed.ncbi.nlm.nih.gov/28292740/)]
18. Litchman ML, Edelman LS, Donaldson GW. Effect of diabetes online community engagement on health indicators: Cross-sectional study. *JMIR Diabetes* 2018 Apr 24;3(2):e8 [FREE Full text] [doi: [10.2196/diabetes.8603](https://doi.org/10.2196/diabetes.8603)] [Medline: [30291079](https://pubmed.ncbi.nlm.nih.gov/30291079/)]
19. Yadav A, Phillips MM, Lundeberg MA, Koehler MJ, Hilden K, Dirkin KH. If a picture is worth a thousand words is video worth a million? Differences in affective and cognitive processing of video and text cases. *J Comput High Educ* 2011 Feb 11;23(1):15-37. [doi: [10.1007/s12528-011-9042-y](https://doi.org/10.1007/s12528-011-9042-y)]
20. Tuong W, Larsen ER, Armstrong AW. Videos to influence: A systematic review of effectiveness of video-based education in modifying health behaviors. *J Behav Med* 2014 Apr;37(2):218-233. [doi: [10.1007/s10865-012-9480-7](https://doi.org/10.1007/s10865-012-9480-7)] [Medline: [23188480](https://pubmed.ncbi.nlm.nih.gov/23188480/)]
21. Kemp S. Digital 2021: Taiwan. DataReportal. 2021 Feb 11. URL: <https://datareportal.com/reports/digital-2021-taiwan> [accessed 2022-03-17]

22. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *Int J Surg* 2012;10(1):28-55 [FREE Full text] [doi: [10.1016/j.ijisu.2011.10.001](https://doi.org/10.1016/j.ijisu.2011.10.001)] [Medline: [22036893](https://pubmed.ncbi.nlm.nih.gov/22036893/)]
23. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007 May;39(2):175-191. [doi: [10.3758/BF03193146](https://doi.org/10.3758/BF03193146)]
24. Drug Safety Program. School of Pharmacy, Taipei Medical University. URL: <http://pharmschool.tmu.edu.tw/activity/index.php?type=20> [accessed 2022-03-17]
25. LINE official account. LINE Biz-Solutions. URL: <https://tw.linebiz.com/service/account-solutions/line-official-account/> [accessed 2022-03-17]
26. Kuo YF. Psychometric Evaluation of the Simplified Diabetes Knowledge Scale and Attitudes Towards Diabetes Scale in Taiwan: Translation and Validation [master's thesis]. Taipei City, Taiwan: Department of Clinical Pharmacy, School of Pharmacy, Taipei Medical University; 2020. URL: <https://hdl.handle.net/11296/yfd67r> [accessed 2022-03-17]
27. Collins G, Mughal S, Barnett A, Fitzgerald J, Lloyd C. Modification and validation of the Revised Diabetes Knowledge Scale. *Diabet Med* 2011 Mar;28(3):306-310. [doi: [10.1111/j.1464-5491.2010.03190.x](https://doi.org/10.1111/j.1464-5491.2010.03190.x)] [Medline: [21309839](https://pubmed.ncbi.nlm.nih.gov/21309839/)]
28. Fitzgerald JT, Davis WK, Connell CM, Hess GE, Funnell MM, Hiss RG. Development and validation of the Diabetes Care Profile. *Eval Health Prof* 1996 Jun 29;19(2):208-230. [doi: [10.1177/016327879601900205](https://doi.org/10.1177/016327879601900205)] [Medline: [10186911](https://pubmed.ncbi.nlm.nih.gov/10186911/)]
29. Hsieh LC. The Relationships Among Disease Knowledge, Social Support and Self-Care Behavior of the Homeless With Type 2 Diabetes [master's thesis]. Taipei City, Taiwan: School of Nursing, National Taipei University of Nursing and Health Sciences; 2015 Jul. URL: <http://irlib.ntunhs.edu.tw/retrieve/5580/103NTCN0563089-001.pdf> [accessed 2022-03-17]
30. Toobert DJ, Hampson SE, Glasgow RE. The summary of diabetes self-care activities measure: Results from 7 studies and a revised scale. *Diabetes Care* 2000 Jul;23(7):943-950. [doi: [10.2337/diacare.23.7.943](https://doi.org/10.2337/diacare.23.7.943)] [Medline: [10895844](https://pubmed.ncbi.nlm.nih.gov/10895844/)]
31. Weiss BD, Mays MZ, Martz W, Castro KM, DeWalt DA, Pignone MP, et al. Quick assessment of literacy in primary care: The newest vital sign. *Ann Fam Med* 2005 Nov;3(6):514-522 [FREE Full text] [doi: [10.1370/afm.405](https://doi.org/10.1370/afm.405)] [Medline: [16338915](https://pubmed.ncbi.nlm.nih.gov/16338915/)]
32. Lin CW. Validation and implication for the Chinese version of the Newest Vital Sign health literacy scale: A case study of pre-service early childhood teachers [article in Chinese]. *J Health Promot Health Educ Contents* 2010 Dec;34:1-31 [FREE Full text]
33. Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? *Int J Methods Psychiatr Res* 2011 Mar 24;20(1):40-49 [FREE Full text] [doi: [10.1002/mpr.329](https://doi.org/10.1002/mpr.329)] [Medline: [21499542](https://pubmed.ncbi.nlm.nih.gov/21499542/)]
34. Zhang L, He X, Shen Y, Yu H, Pan J, Zhu W, et al. Effectiveness of smartphone app-based interactive management on glycemic control in Chinese patients with poorly controlled diabetes: Randomized controlled trial. *J Med Internet Res* 2019 Dec 09;21(12):e15401 [FREE Full text] [doi: [10.2196/15401](https://doi.org/10.2196/15401)] [Medline: [31815677](https://pubmed.ncbi.nlm.nih.gov/31815677/)]
35. Lotto M, Strieder AP, Ayala Aguirre PE, Oliveira TM, Andrade Moreira Machado MA, Rios D, et al. Parental-oriented educational mobile messages to aid in the control of early childhood caries in low socioeconomic children: A randomized controlled trial. *J Dent* 2020 Oct;101:103456 [FREE Full text] [doi: [10.1016/j.jdent.2020.103456](https://doi.org/10.1016/j.jdent.2020.103456)] [Medline: [32827598](https://pubmed.ncbi.nlm.nih.gov/32827598/)]
36. Goodarzi M, Ebrahimzadeh I, Rabi A, Saedipoor B, Jafarabadi MA. Impact of distance education via mobile phone text messaging on knowledge, attitude, practice and self efficacy of patients with type 2 diabetes mellitus in Iran. *J Diabetes Metab Disord* 2012 Aug 31;11(1):10 [FREE Full text] [doi: [10.1186/2251-6581-11-10](https://doi.org/10.1186/2251-6581-11-10)] [Medline: [23497632](https://pubmed.ncbi.nlm.nih.gov/23497632/)]
37. Devi GP. Effectiveness of digital media education on the attitude of the diabetic patients. *Int J Curr Microbiol Appl Sci* 2020 Sep 10;9(9):1116-1123 [FREE Full text] [doi: [10.20546/ijcmas.2020.909.139](https://doi.org/10.20546/ijcmas.2020.909.139)]
38. Belsti Y, Akalu Y, Anmut Y. Attitude, practice and its associated factors towards diabetes complications among type 2 diabetic patients at Addis Zemen District hospital, Northwest Ethiopia. *BMC Public Health* 2020 May 26;20(1):785 [FREE Full text] [doi: [10.1186/s12889-020-08953-6](https://doi.org/10.1186/s12889-020-08953-6)] [Medline: [32456637](https://pubmed.ncbi.nlm.nih.gov/32456637/)]
39. Rahaman KS, Majdzadeh R, Holakouie Naieni K, Raza O. Knowledge, attitude and practices (KAP) regarding chronic complications of diabetes among patients with type 2 diabetes in Dhaka. *Int J Endocrinol Metab* 2017 Jul 30;15(3):e12555 [FREE Full text] [doi: [10.5812/ijem.12555](https://doi.org/10.5812/ijem.12555)] [Medline: [29201069](https://pubmed.ncbi.nlm.nih.gov/29201069/)]
40. Guze PA. Using technology to meet the challenges of medical education. *Trans Am Clin Climatol Assoc* 2015;126:260-270 [FREE Full text] [Medline: [26330687](https://pubmed.ncbi.nlm.nih.gov/26330687/)]
41. Taylor D, Grant J, Hamdy H, Grant L, Marei H, Venkatramana M. Transformation to learning from a distance. *MedEdPublish* 2020 Apr 23;9(1):76 [FREE Full text] [doi: [10.15694/mep.2020.000076.1](https://doi.org/10.15694/mep.2020.000076.1)]
42. 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](https://doi.org/10.1001/jama.2020.5227)] [Medline: [32232420](https://pubmed.ncbi.nlm.nih.gov/32232420/)]
43. Bhaskar S, Bradley S, Chattu VK, Adishes A, Nurtazina A, Kyrkybayeva S, et al. Telemedicine as the new outpatient clinic gone digital: Position paper from the Pandemic Health System REsilience PROGRAM (REPROGRAM) International Consortium (Part 2). *Front Public Health* 2020;8:410 [FREE Full text] [doi: [10.3389/fpubh.2020.00410](https://doi.org/10.3389/fpubh.2020.00410)] [Medline: [33014958](https://pubmed.ncbi.nlm.nih.gov/33014958/)]
44. Alromaihi D, Alamuddin N, George S. Sustainable diabetes care services during COVID-19 pandemic. *Diabetes Res Clin Pract* 2020 Aug;166:108298 [FREE Full text] [doi: [10.1016/j.diabres.2020.108298](https://doi.org/10.1016/j.diabres.2020.108298)] [Medline: [32623031](https://pubmed.ncbi.nlm.nih.gov/32623031/)]
45. Hare N, Bansal P, Bajowala SS, Abramson SL, Chervinskiy S, Corriel R, et al. Work group report: COVID-19: Unmasking Telemedicine. *J Allergy Clin Immunol Pract* 2020 Sep;8(8):2461-2473.e3 [FREE Full text] [doi: [10.1016/j.jaip.2020.06.038](https://doi.org/10.1016/j.jaip.2020.06.038)] [Medline: [32603900](https://pubmed.ncbi.nlm.nih.gov/32603900/)]

46. Sørensen K, Pelikan JM, Röthlin F, Ganahl K, Slonska Z, Doyle G, HLS-EU Consortium. Health literacy in Europe: Comparative results of the European Health Literacy Survey (HLS-EU). *Eur J Public Health* 2015 Dec 05;25(6):1053-1058 [FREE Full text] [doi: [10.1093/eurpub/ckv043](https://doi.org/10.1093/eurpub/ckv043)] [Medline: [25843827](https://pubmed.ncbi.nlm.nih.gov/25843827/)]
47. Stewart CC, Yu L, Lamar M, Wilson RS, Bennett DA, Boyle PA. Associations of health and financial literacy with mortality in advanced age. *Aging Clin Exp Res* 2020 May 04;32(5):951-957 [FREE Full text] [doi: [10.1007/s40520-019-01259-7](https://doi.org/10.1007/s40520-019-01259-7)] [Medline: [31273677](https://pubmed.ncbi.nlm.nih.gov/31273677/)]
48. Singh-Manoux A, Kivimaki M, Glymour MM, Elbaz A, Berr C, Ebmeier KP, et al. Timing of onset of cognitive decline: Results from Whitehall II prospective cohort study. *BMJ* 2012 Jan 05;344:d7622 [FREE Full text] [doi: [10.1136/bmj.d7622](https://doi.org/10.1136/bmj.d7622)] [Medline: [22223828](https://pubmed.ncbi.nlm.nih.gov/22223828/)]
49. Abdullah A, Liew SM, Salim H, Ng CJ, Chinna K. Prevalence of limited health literacy among patients with type 2 diabetes mellitus: A systematic review. *PLoS One* 2019;14(5):e0216402 [FREE Full text] [doi: [10.1371/journal.pone.0216402](https://doi.org/10.1371/journal.pone.0216402)] [Medline: [31063470](https://pubmed.ncbi.nlm.nih.gov/31063470/)]
50. Tseng HM, Liao SF, Wen YP, Chuang YJ. Adaptation and validation of a measure of health literacy in Taiwan: The Newest Vital Sign. *Biomed J* 2018 Aug;41(4):273-278 [FREE Full text] [doi: [10.1016/j.bj.2018.07.001](https://doi.org/10.1016/j.bj.2018.07.001)] [Medline: [30348271](https://pubmed.ncbi.nlm.nih.gov/30348271/)]
51. Huang ST, Chuang HY, Chang WH, Huang MY. Explore patient-specific needs for mHealth solutions in Taiwan: A mixed study. *Health Technol* 2018;2(4):1-6 [FREE Full text] [doi: [10.21037/ht.2018.03.01](https://doi.org/10.21037/ht.2018.03.01)]
52. CDCES Fact Sheet. Arlington Heights, IL: Certification Board for Diabetes Care and Education; 2020. URL: https://www.cbdcce.org/documents/20123/66178/CDCES_factsheet_4c_WEB.pdf/3a1c503d-cbde-72af-b50f-b2385e4e6ae4?t=1587662753440 [accessed 2022-03-17]
53. Kim SH, Utz S. Effectiveness of a social media-based, health literacy-sensitive diabetes self-management intervention: A randomized controlled trial. *J Nurs Scholarsh* 2019 Nov 17;51(6):661-669. [doi: [10.1111/jnu.12521](https://doi.org/10.1111/jnu.12521)] [Medline: [31622033](https://pubmed.ncbi.nlm.nih.gov/31622033/)]

Abbreviations

AADE7: American Association of Diabetes Educators 7 Self-Care Behaviors

CDE: certified diabetes educator

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

DCP-ATDS: Diabetes Care Profile–Attitudes Toward Diabetes Scales

FAQ: frequently asked questions

HbA_{1c}: glycated hemoglobin

mHealth: mobile health

NVS: Newest Vital Sign

OR: odds ratio

RCT: randomized controlled trial

SDKS: Simplified Diabetes Knowledge Scale

SDSCA: Summary of Diabetes Self-Care Activities

TMU: Taipei Medical University

TMU-LOVE: Taipei Medical University–LINE Oriented Video Education

Edited by A Mavragani; submitted 26.06.21; peer-reviewed by M Lotto, CY Lin; comments to author 19.07.21; revised version received 03.08.21; accepted 04.02.22; published 23.03.22.

Please cite as:

Leong CM, Lee TI, Chien YM, Kuo LN, Kuo YF, Chen HY

Social Media–Delivered Patient Education to Enhance Self-management and Attitudes of Patients with Type 2 Diabetes During the COVID-19 Pandemic: Randomized Controlled Trial

J Med Internet Res 2022;24(3):e31449

URL: <https://www.jmir.org/2022/3/e31449>

doi: [10.2196/31449](https://doi.org/10.2196/31449)

PMID: [35319478](https://pubmed.ncbi.nlm.nih.gov/35319478/)

©Cheng Man Leong, Ting-I Lee, Yu-Mei Chien, Li-Na Kuo, Yu-Feng Kuo, Hsiang-Yin Chen. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 23.03.2022. 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.

Original Paper

Review of Mobile Apps for Women With Anxiety in Pregnancy: Maternity Care Professionals' Guide to Locating and Assessing Anxiety Apps

Kerry Evans^{1*}, BSc, MA, PhD; Jasper Donelan^{2*}, PhD; Stefan Rennick-Egglestone^{3*}, PhD, MA; Serena Cox^{1*}, BSocSci; Yvonne Kuipers^{4*}, PhD

¹School of Health Sciences, University of Nottingham, Nottingham, United Kingdom

²Digital Research, University of Nottingham, Nottingham, United Kingdom

³School of Health Sciences, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

⁴Edinburgh Napier University, School of Health and Social Care, Edinburgh, United Kingdom

* all authors contributed equally

Corresponding Author:

Kerry Evans, BSc, MA, PhD

School of Health Sciences

University of Nottingham

Medical School

Queen's Medical Centre

Nottingham, NG7 2HA

United Kingdom

Phone: 44 115 95 15559

Email: kerry.evans1@nottingham.ac.uk

Abstract

Background: Mental health and pregnancy apps are widely available and have the potential to improve health outcomes and enhance women's experience of pregnancy. Women frequently access digital information throughout their pregnancy. However, health care providers and women have little information to guide them toward potentially helpful or effective apps.

Objective: This review aimed to evaluate a methodology for systematically searching and reviewing commercially available apps that support pregnant women with symptoms of anxiety in order to assist maternity care professionals in identifying resources that they could recommend for these women.

Methods: A stepwise systematic approach was used to identify, select, describe, and assess the most popular and highly user-rated apps available in the United Kingdom from January to March 2021. This included developing a script-based search strategy and search process, writing evaluation criteria, and conducting a narrative description and evaluation of the selected apps.

Results: Useful search terms were identified, which included nonclinical, aspirational, and problem-based phrases. There were 39 apps selected for inclusion in the review. No apps specifically targeted women with anxiety in pregnancy. Of the 39 apps included in the review, 33 (85%) focused solely on mind-body techniques to promote relaxation, stress reduction, and psychological well-being. Only 8 of the 39 (21%) apps included in the review reported that health care professionals had contributed to app development and only 1/39 (3%) provided empirical evidence on the effectiveness and acceptability of the app. The top 12/39 (31%) apps were evaluated by 2 independent reviewers using the developed criteria and scores. There was a small negative correlation between the reviewers' scores and app user rating scores, with higher user rating scores associated with lower reviewer scores.

Conclusions: App developers, publishers, and maternity care professionals should seek advice from women with lived experience of anxiety symptoms in pregnancy to locate, promote, and optimize the visibility of apps for pregnant women. There is a lack of resources that provide coping strategies based on current evidence for the treatment of anxiety in pregnancy. Maternity care providers are limited in their ability to locate and recommend acceptable and trustworthy apps because of the lack of information on the evidence base, development, and testing of apps. Maternity care professionals and women need access to libraries of trusted apps that have been evaluated against relevant and established criteria.

KEYWORDS

anxiety; pregnancy; antenatal; mobile applications; digital interventions; mHealth; mobile app; psychological well-being; maternity; evaluation; quality assessment

Introduction

Many pregnant women experience symptoms of anxiety; the prevalence of antenatal anxiety symptoms has been reported to be 13%-23% [1-3]. During the COVID-19 pandemic, the number of women with symptoms of anxiety in pregnancy has increased due to women's concerns about virus transmission, accessing care, and social support [4,5]. Anxiety symptoms in pregnancy usually have similar affective and cognitive attributes to anxiety symptoms at other times [6], although concerns related to pregnancy may present as the predominant feature. Mild anxiety in pregnancy may be normal to prepare women for motherhood and protecting the fetus [6,7]. Anxiety symptoms become problematic when a significant amount of a woman's time is consumed, when women are unable to redirect their focus to other tasks, or when everyday life and relationships are affected [6,8,9]. Antenatal anxiety is reported to be associated with postpartum depression, greater use of interventions during labor, reduced rates of breastfeeding, prematurity, and preterm birth [10-12].

The provision of web-based advice via mobile phones and the internet has been suggested to help reduce anxiety and stress in pregnant women [4,13]. Digitally delivered interventions are potential solutions to overcoming barriers to access treatment for perinatal mental health disorders. Interventions can be delivered as unguided resources to support or replace patient-provider interactions or as guided interventions that may include live interactions over telephone or video or contact with therapists using digital messaging [14]. Increasing midwives' and maternity care providers' awareness of digitally delivered information and supportive interventions could assist in signposting pregnant women to effective resources [15]. During pregnancy, women frequently access digital information [16], although information accessed through web-based sources is rarely discussed with health care providers and the providers themselves may be unaware of web-based information and its accuracy [17]. Mobile apps have the potential to positively influence health behavior and health outcomes [18,19] and enhance women's experience of maternity care and pregnancy [20]. The availability of mobile apps has increased significantly over the past few years. There are currently more than 400,000 health apps available from Google Play [21] and the App Store [22]; however, there are fewer than 10,000 downloads for many of these apps and 25% are never used after installation [18].

Mental health and pregnancy apps are widely available; however, health care providers and women have little information on which apps may be helpful or which to avoid because these are ineffective or have potentially harmful content [23]. Platforms such as the NHS Apps Library [24] have only recently been developed to assist patients and the public in finding trusted health and well-being apps and include general pregnancy and mental health apps [25]. User ratings presented

by the app stores, on the other hand, do not provide a measure of clinical appropriateness, safety, or efficacy, and the availability of clinical data to guide app recommendations is poor [23]. Moreover, the health app industry is commercially dominated and lacks regulation [16].

Health apps for use in pregnancy or to support individuals with anxiety symptoms have been assessed as having poor quality [26-28], lacking evidence-based content [29], and being ineffective or potentially harmful [30]. Other reviews of pregnancy apps have reported that apps contained little or no pregnancy-specific information [26] and contained information that was potentially harmful for pregnant women [31]. Health professionals and app users have reported a preference for using pregnancy apps that are relevant to their local health care context and come from a trusted source. There is a need for greater health professional engagement in app development and increased awareness of and guidance for use of these resources [16].

Several methodologies have been proposed for evaluating the quality of apps [23,25], but the validity of app rating measures is not yet established [18] and standardized measures with high interrater reliability are required. Nouri et al [18] conducted a systematic review of existing health app assessment tools (N=23). In total, there were 38 main classes of assessment criteria, which the reviewers arranged into 7 main criteria: Design, Information/Content, Usability, Functionality, Ethical Issues, Security and Privacy, and User-perceived value. Powell et al [23] evaluated the interrater reliability of existing app quality measures, but only criteria for "App interactivity" and "feedback" reached the threshold for agreement, with items related to security and privacy, number of ratings, research base, authorship, attribution, and product advisory support reaching near threshold levels of agreement (α levels of .5 or more). Items with low interrater reliability (α .3 or less) included more subjective measures, including perceived and claimed effectiveness, ease of use, errors, and performance issues. The authors state that even more objective measures can be missed by a reviewer and suggest that the evidence base of the app may be a more reliable indicator of effectiveness. Therefore, clinicians need to review apps personally before making recommendations, discuss apps with colleagues and service users, and apply clinical judgment [23].

The purpose of this review was to evaluate a methodology to systematically search and review commercially available apps to support pregnant women with symptoms of anxiety. The review focused on identifying app resources that could be used to complement maternity and perinatal mental health care and identifying methods of searching and evaluation that can be adopted by maternity care professionals with limited time and resources.

Methods

This review adopted a stepwise systematic approach to identify, select, describe, and assess the most popular and highly user-rated apps available in the United Kingdom from January to March 2021. The review team included individuals with expertise in maternity care and research, mental health care, and digital research and a lived experience of maternity. Ethics approval was not required for this review because primary data were not collected.

Developing an App Search Strategy

The first stage involved identifying key search words and terms (Table 1). Research and academic keywords and Medical Subject Headings terms (eg, Anxiety AND Pregnancy) were piloted on Android (Google Play, Google) and iOS (App Store, Apple). These search terms led to limited results and did not locate preidentified “marker” apps (Mind the Bump [32] and Headspace [33]). The Headspace and Mind the Bump apps were selected as marker apps because they are recommended by the National Health Service or established mental health charities. General phrases were brainstormed by the review team, considering terms that women may use to search for resources

as well as descriptive and marketing terms that app developers might employ. The search strategy also considered the work by Wexler et al [34] in a study related to our own. Wexler et al [34] analyzed data from pregnancy social media forums. They found that the frequency of word appearance in relation to other words revealed clusters of words that have a high probability of appearing together around certain topics. Therefore, word clusters for anxiety about pregnancy and labor were included in the search terms. Members of an established public involvement group also contributed by identifying words they would use to locate information and resources. Service users highlighted that women may also search for specific pregnancy-related aspects that cause anxiety (eg, miscarriage, giving birth) or use words to describe how they would like to feel (eg, calm, stress-free, relaxed). Again, these were included in the key search words and terms for the present review.

Apps were included in the review if they were available in English, aimed at or referred to use in pregnancy, and included advice, information, exercises, or techniques to improve symptoms of anxiety, worry, fear, or stress. Exclusion criteria were anxiety apps that did not reference pregnancy in the app information and pregnancy apps that did not reference anxiety, worry, stress, fear, or emotional or mental health.

Table 1. Developing search terms for Google Play and the App Store.

Research literature searching keywords	Web-based words and phrases
Pregnancy, antenatal, perinatal, childbearing	Pregnant, pregnancy, motherhood, mother (mum, mom, mama, mamma), baby, birth (childbirth), miscarriage, movements (fetal/baby movements), labor, maternity
Anxiety	Anxious (anxiety), worry (worries), concerns; stress, distress; ear, panic, scared, nervous; mind, emotion, thoughts, mood; mental health; therapy (CBT ^a); relax, calm; cope (coping); help, care, relief, cure; wellbeing

^aCBT: cognitive behavioral therapy.

Search Process

To identify efficient and effective search terms that could be used by maternity care professionals, previously developed scripts were used to search the UK version of Google Play and the App Store. The search was conducted from January to March 2021 using the script-based approach developed by Stawarz et al [35]. The scripts were modified using alternative combinations of the suggested keywords. Each combination of keywords and the number of apps located are shown in Multimedia Appendix 1. Search results included the app name, short description, rating, and developer’s details. After script-based searching, a simple web-based keyword search of Google Play and the App Store was performed to compare the search results with those of script-based searching and to locate additional apps. Changing the country code (ie, United Kingdom, United States) did not appear to impact the search result. Changing the word order resulted in different search results on Google Play but not on the App Store. Manual screening of the app search results was completed by evaluating the app title and description against the inclusion and exclusion criteria.

Identifying Evaluation Criteria for the Apps

Evaluation criteria that considered the findings from existing pregnancy and mental health app reviews were selected. The

criteria were therapeutic or supportive content (reflective of the evidence base or clinical standards) [29,35,36], relevance for a pregnant population [26,31,37], and compliance with existing app quality measures (Multimedia Appendix 2). The following therapeutic or supportive content was evaluated against review-level evidence and clinical recommendations for perinatal and common mental health disorders [38-41]:

1. Psychoeducation: Aids in the identification of anxiety disorders and improves understanding and treatment options for anxiety.
2. Low-intensity psychological interventions including individual nonfacilitated, guided self-help or psychoeducational group. Includes written or electronic materials based on the treatment principles of cognitive behavioral therapy (CBT).
3. Mind-body interventions such as yoga or hypnotherapy.
4. Referral and signposting to services for women requiring specialist diagnosis and care.

It is important to consider whether the current evidence for an app is sufficient or relevant for a particular population [37]. Therefore, the following questions were developed by the reviewers to identify the availability of information that women could search for [42].

1. How do I know if I have anxiety?
2. How do I know the severity of my anxiety symptoms?
3. Where can I go for help for my anxiety in pregnancy?
4. How can I access specialist help for my anxiety in pregnancy?
5. What are the treatment options for anxiety during pregnancy?
6. What can I do to help my anxiety symptoms?

Based on the review by Nouri et al [18], quality criteria were selected to meet the review objectives. Subjective measures of app quality have been reported as having low interrater reliability [23]. Quality criteria were therefore selected that focused on objective measures and minimized the number of subjective responses. Efforts to reduce the subjective measures also considered end-user preference. Criteria identified by health care professionals as important may not reflect women’s needs and preferences, as different individuals may find a particular app more or less useful depending on their personal needs, circumstances, experiences, and education level [25,43]. User ratings as reported in Google Play and the App Store were also presented alongside reviewers’ evaluation scores for comparison.

Analysis

Descriptive data were collected (where available) for all apps included in the review. This included the app developer and country, number of downloads, number of ratings, rating score, version number, costs, and date of the last update. Narrative description and evaluation of the highest scoring, free to download, and most popular apps were completed [29]. Apps were identified based on the app download numbers and user rating (0-5 stars) provided by Google Play and the App Store: more than 10,000 installs, a star rating of 3.5 or higher, and free to download in Google Play; and more than 10 ratings, a star rating of 3.5 or higher, and free to download in the App Store.

This resulted in the selection of 12 apps that were then evaluated by 2 independent reviewers using the criteria presented in [Multimedia Appendix 2](#). The extent to which the app addressed the criteria was scored on a 2-point scale (0, information is absent/negative response; 1, information is present/positive response). The overall score was calculated and interpreted with caution because of the lack of validity and reliability of app quality measures [23]. Scores were used to assist in interpreting the review findings and to generate discussion on the mechanisms for selecting and recommending potentially beneficial apps ([Table 2](#)).

Table 2. Quality scores of the highest user rated and most popular apps in the review.

App name	Maternity and mental health policy and evidence base score (out of 6)	General app features quality score (out of 8)	Combined score (out of 14)	User rating score from Google play/App store (out of 5)
Antenatal Yoga, Meditation + Education (YogiBirth)	1	3	4	4.3/4.7
Baby Buddy	4	7	11	2.8/4
Carry: Pregnancy Workouts	5	4	9	—/4.7
Hypnobirthing: Calm Birth	1	3	4	—/4.4
Hypnobirthing - Pregnancy, Music & Tracker	1	3	4	3.4/—
IHypnobirth	1	4	5	—/3.8
Keleya: Pregnancy Fitness & Tracker + Baby Due Date	1	5	6	4.1/—
Music for Pregnancy Relaxation	0	3	3	4.2/—
Pregnancy Care Tips	0	1	1	3.6/—
Pregnancy Music Collection 200	0	2	2	4.4/—
Pregnancy Yoga Exercises	1	4	5	4.0/—
Pregnancy Yoga Exercises – Prenatal Yoga	1	3	4	3.8/—

Results

Apps Located and Included in the Review

In total, 1391 apps were located, and 1337 (96%) apps were subsequently excluded based on the title, description, and duplication ([Figure 1](#)). A total of 39/1391 (3%) apps were included in the review ([Multimedia Appendix 3](#)).

No apps specifically targeted women with anxiety in pregnancy. Various apps focused on providing well-being support during

pregnancy and included stress relief, relaxation, and mental health advice as part of a general approach to well-being in pregnancy. Of the 39 apps included in the initial review, 33 (85%) focused solely on mind-body techniques to promote relaxation, stress reduction, and psychological well-being. Mind-body techniques included relaxation, mindfulness, hypnosis, yoga, positive affirmations, and meditation. A total of 3/39 (8%) apps provided informational support, of which 1 (33%) focused on psychoeducation and 2 (67%) were multicomponent and provided information support, cognitive,

or mind-body techniques. Only 8 of the 39 (21%) apps included in the review reported that health care professionals had contributed to app development (Figure 2). Only 1/39 (3%) app (Baby Buddy) provided empirical evidence on the effectiveness and acceptability of the app.

There was a small negative correlation between the reviewers' scores and the user rating scores ($r=-0.27$; 12/39, 31%; $P=.39$) with higher user rating scores associated with lower reviewer scores.

Figure 1. Adapted PRISMA diagram: anxiety apps for pregnant women. PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

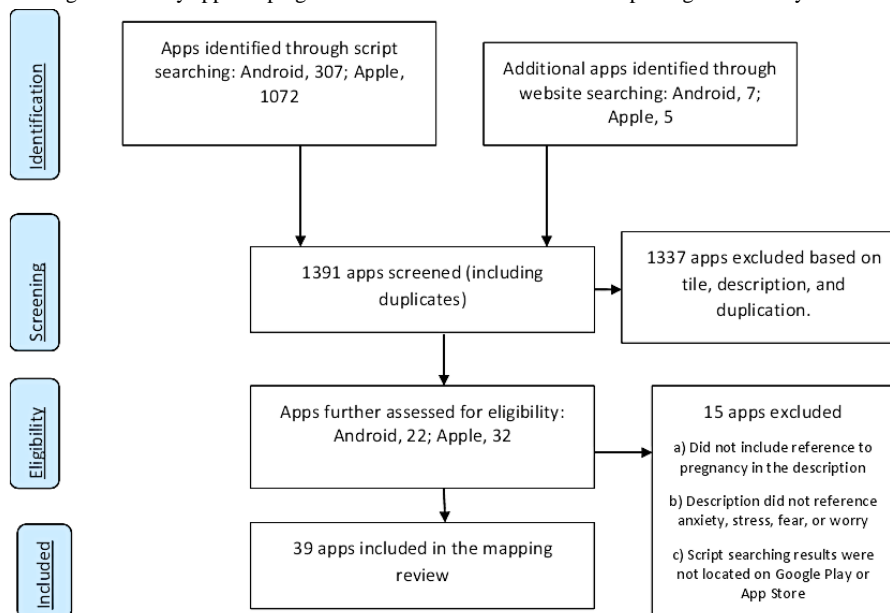


Figure 2. Yoga, relaxation and hypnobirthing apps.

	YogiBirth Pregnancy Yoga, Meditation + Education	Hypnobirthing –Pregnancy Music and Tracker	Pregnancy Yoga Exercises	Pregnancy Yoga Exercises – Prenatal Yoga	Carry: Pregnancy Workouts	Hypnobirthing: MoM Pregnancy	iHypnobirth – lite
Android	✓	✓	✓	✓			
iOS		✓			✓	✓	✓
Additional content changes	✓	✓				✓	✓
Yoga	✓		✓	✓	✓		
Meditation / relaxation	✓	✓			✓	✓	✓
Positive affirmations		✓				✓	
Hypnobirthing		✓				✓	✓
Provision or signposting to free evidence- based information or strategies					✓		
Content development described	Midwife and antenatal yoga teacher				Yoga teacher, Doula and Physical Therapist		Psychologist & hypnatal practitioner
Updated within 12 months of the review	Within 8 months of the review	Within 2 months of the review	Within 1 months of the review	Within 7 months of the review			

Description of the Apps Included in the Evaluation

Yoga, Relaxation, Meditation, and Hypnobirthing Apps

Of the 12 apps, 7 (58%) provided elements of relaxation, hypnobirthing, yoga, and relaxation (Figure 2) [44-50]. Of the

7 apps, 4 (57%) provided exercises to improve mood, anxiety coping, or mental well-being [44-47], although no supporting evidence or information about anxiety or mental well-being was provided. Only 1/7 (14%) app [48] included information about depression and anxiety in pregnancy and the postnatal period

via links to national health care websites and helplines for factsheets.

Music Apps

Of the 12 apps, 2 (17%) provided music for relaxation in pregnancy and were very similar in content and appearance: Music for Pregnancy Relaxation [51] and Pregnancy Music Collection [52]. Updates had not been completed within 12 months of the review. Both apps reported that music could reduce stress and anxiety in pregnancy and support fetal brain development. No empirical evidence was provided to support the claims. Both apps provided recordings of classical music or soothing sounds through a simple interface.

Informative Apps

Pregnancy Care Tips [53] was available on Android and iOS devices, although it had not been updated within 5 years of the review. At the time of publication, the app was no longer available to download. The app provided written “tips” and information on pregnancy symptoms, diet, exercise, and maternity care. The information appeared to have been translated into English and certain phrases and terms were difficult to understand and not reflective of UK maternity care procedures. No information was provided on the development of the app, professional input, or supporting evidence.

Keleya: Pregnancy Fitness & Tracker + Baby Due Date [54] was available on Android and iOS devices. Keleya was promoted as an “All-in-One App” for pregnancy and contains yoga exercises, nutritional advice, and information on the progress and symptoms of pregnancy. Content was provided in written and audio material, which could be personalized by responding to user input and detailing stage of pregnancy, goals, and symptoms. The app reported reducing anxiety through meditation exercises. No empirical evidence was provided, and there was no information on app development or health care professional input. Free content was very limited, and additional features such as social media community access, information, and exercises required a subscription fee. Information on anxiety in pregnancy was limited to user input to track anxiety symptoms with no further information, support, or signposting available.

Baby Buddy [55] is a National Health Service–endorsed app from the charity Best Beginnings. The app was updated within 1 month of the review and is free on Android and IOS devices. A website link from the app provided links to evaluation and impact studies within UK maternity care institutions. The app reported guiding women through pregnancy by providing information through written and video materials. Baby Buddy provided numerous video clips on the topic of anxiety and depression, although there were no specific resources to help women develop coping strategies and techniques to manage symptoms of anxiety. The information included an overview of anxiety disorders, CBT approaches, how or when to seek help with mental health, and the benefits of peer support. Baby Buddy reported that the text and video clips were quality-assured by professional Royal Colleges and other health organizations. The app was interactive and enabled women to make notes of questions to ask their midwife and input information to access

personalized content. Signposting was provided for local maternity services and support groups.

Discussion

Locating and Evaluating Apps

The purpose of this review to evaluate a methodology to systematically search for and review commercially available apps was addressed by identifying useful search terms that were nonclinical and included aspirational phrases as suggested by the study service user group. The terms included “calm,” “relax,” “relief,” “cope,” and “well-being” as well as some problem-based phrases such as “stress,” “fear,” and “anxiety.” However, no clear recommendations were identified for the searching of apps to help maternity care professionals identify useful resources.

The evaluation components developed for the review provided a useful framework for maternity care professionals to assess the therapeutic or supportive content in the context of pregnancy. The quality criteria focused on more objective measures and aimed to minimize the number of subjective responses. However, the binary coding (1, information/component present; 0, information/component absent) was not amenable to capture criteria that were partly addressed by the app or where particular types of content were presented differently within the app [42,56]. The quality criteria have been revised in response to the findings and are presented in [Multimedia Appendix 4](#). Further studies to evaluate the reliability of the criteria are required [18].

Many of the apps that purported to reduce anxiety symptoms did not include any link to peer-reviewed literature or the evidence base. Other reviews have reported the lack of provision of evidence-based app content [28,35,36,56]. Although web-based mindfulness and CBT approaches have been reported as effective in reducing anxiety and other mental health concerns in perinatal populations [57,58], evidence-based information was not presented in the apps. The lack of clearly defined content with links to the evidence base may hinder the ability of maternity care professionals to determine the quality of the app [56]. However, this review highlighted that only 1 of the 7 (14%) most popular and highly rated mind-body apps provided any rationale for the approach or links to the evidence base.

The lack of correlation between the presence of evidence-based information or strategies and the popularity of mental health apps has been highlighted in recent reviews [35,56,59]. This suggests that features other than the provision of evidence-based information are important to and valued by women [35]. To further highlight this point, this review has demonstrated that the quality evaluation scores assigned by the review team (health care professionals and researchers) did not reflect user ratings (displayed by the app platforms). Because of the lack of information on the development of the apps, it is difficult to know whether women’s views were accessed or whether the design and content of the app reflected women’s needs.

User experience and engagement are important factors in the overall effectiveness of apps [35]. Stawarz et al [35] identified that interactive features and customization were important in improving user engagement with CBT for depression. Positivity

(ability to capture positive and negative thoughts) along with privacy, security, and trust was associated with improved user experience. For this review, engagement with service users also highlighted that women may wish to use “positive” and aspirational words to describe how they would like to feel (eg, calm, stress-free, relaxed) when they search for potentially useful apps. Researchers should aim to make mobile health interventions with evidence-based treatments attractive and accessible and focus on user desirability and experience early in the design phase [59]. A combination of approaches to maximize user engagement and experience with evidence-based strategies is required to deliver potentially effective app-based strategies to support women with anxiety in pregnancy.

Pregnant women may benefit from remotely delivered interventions to help them cope with symptoms of anxiety if they are provided with web-based contact with a health care professional or peer community and may be more motivated to complete interventions that are perceived as relevant or tailored to their needs and situations [60]. Only 2 of the 39 (5%) apps included in the review provided a psychological therapeutic approach, with most apps providing mind-body techniques and exercises to support women’s general well-being throughout pregnancy, labor, and birth. Only 1/39 (3%) app was endorsed by a health care organization (Baby Buddy). Although it did not provide any therapeutic content, this app did provide information and signposting. Apps that offered coping strategies for women with symptoms of anxiety had very little information about how or when to seek help or signposting to supportive services.

The apps included in this review were promoted as either general pregnancy well-being apps (including diet, exercise, fetal growth and well-being, labor, and birth) or general relaxation apps. This review did not locate any apps that were solely focused on anxiety symptoms in pregnancy. A systematic review of perinatal web-based psychological treatments for clinical levels of maternal anxiety and depression [61] also did not locate any interventions targeted to the reduction of anxiety disorders or comorbid depression and anxiety. Although no interventions were tailored to anxiety or recruited women with a diagnosed anxiety disorder, the pooled analysis demonstrated medium and significant group differences favoring web-based interventions over control conditions for anxiety outcome measures. The authors of this previous systematic review recommend that interventions be specifically developed for this neglected area of perinatal mental health care.

The perceived reliability and trustworthiness of web-based pregnancy information have been reported to increase in women when the resource is regularly updated and when it is recommended by a health care professional [62]. Involvement of the health care provider has been reported to help individuals understand what apps can and cannot do [29]. Evaluation of app quality by health care professionals should not be a substitute for women’s preferences, usability, and other personal factors necessary for selecting an app [29]. Zelmer et al [63] sought consensus from a broad group of stakeholders on guiding principles and criteria for a framework to assess e-mental health

apps. The resulting principles are similar to the criteria developed for this review across categories of security, usability, evidence base, and functionality. It would not be possible for most maternity care professionals to assess the fast-growing health apps market. Access to technologies, curated compilations, or libraries of effective apps would, however, help maternity care professionals feel more confident in recommending and signposting women to potentially beneficial resources [25,64]. Platforms such as the NHS Apps Library have started to develop such resources, inviting developers to submit apps for assessment of regulatory, clinical, security, and technological criteria [65].

Study Strengths and Limitations

For this review, potentially useful criteria for rating the quality of the content and the function of apps were suggested to support women with symptoms of anxiety in pregnancy. The criteria were based on previous research, and validity and reliability testing was not conducted nor was it within the remit of the study. The security and privacy policies of each app were not fully scrutinized, although an assessment of whether security and privacy policies were reported in the app information was included. The American Psychiatric Association has developed a framework for app evaluation [37], which begins with the assessment of compliance with safety and privacy criteria. These criteria must be met before the evaluation continues to assess benefit and efficacy, engagement, and data sharing. Simply checking for the existence (or absence) of a privacy policy will help identify questionable apps [37].

Conclusions

Locating potentially useful apps is not a straightforward process and requires a different approach to that used in traditional academic search. Keywords that reflect women’s search queries and that can help women and maternity care providers navigate app libraries need to be developed. App developers, publishers, and maternity care professionals should seek advice from women with lived experience of anxiety symptoms in pregnancy to locate, promote, and optimize the visibility of apps for a diverse population of pregnant women. This review did not locate any resources that provided coping strategies or therapeutic approaches for anxiety that were based on the current evidence base for the treatment of anxiety in pregnancy. The rationale, development, and testing of apps included in this review were underreported, which may hinder the ability of maternity care providers to easily locate useful, acceptable, and trustworthy resources. Potentially useful quality criteria have been presented, which require further development and testing. Maternity care professionals should be aware that features of apps other than the provision of evidence-based information and approaches are important to app users. Features such as interactivity and customization may improve user engagement, and positive framing using aspirational statements may attract women with symptoms of anxiety when they look to select an app resource. Maternity care professionals and women would benefit from access to libraries of trusted apps that have been evaluated against relevant and standardized criteria.

Acknowledgments

The authors would like to acknowledge Dr Katarzyna Stawarz from the University of Cardiff for advice and support with developing the script searches. The study was completed as part of the HEE/NIHR Clinical Lectureship award undertaken by KE. SRE received funding by the National Institute for Health Research (NIHR) through its Programme Grants for Applied Research Programme (Programme Grants for Applied Research, Personal experience as a recovery resource in psychosis: Narrative Experiences Online (NEON) Programme, RP-PG-0615-20016). YK is a management committee member of COST action CA18211: DEVoTION: Perinatal Mental Health and Birth-Related Trauma: Maximising best practice and optimal outcomes. SC received funding by the NIHR Applied Research Collaboration East Midlands (ARC EM). The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Keyword results in the search strategy.

[[DOCX File , 16 KB - jmir_v24i3e31831_app1.docx](#)]

Multimedia Appendix 2

Suggested criteria for evaluating the quality of apps for women with anxiety in pregnancy.

[[DOCX File , 15 KB - jmir_v24i3e31831_app2.docx](#)]

Multimedia Appendix 3

Included apps from Google Play and App Store (information accessed March 4, 2021).

[[DOCX File , 23 KB - jmir_v24i3e31831_app3.docx](#)]

Multimedia Appendix 4

Evaluation criteria for pregnancy anxiety apps (adapted from Van Singer et al [42] and Nouri et al [18]).

[[DOCX File , 15 KB - jmir_v24i3e31831_app4.docx](#)]

References

1. Mitchell-Box K, Braun KL. Fathers' thoughts on breastfeeding and implications for a theory-based intervention. *J Obstet Gynecol Neonatal Nurs* 2012;41(6):E41-E50. [doi: [10.1111/j.1552-6909.2012.01399.x](https://doi.org/10.1111/j.1552-6909.2012.01399.x)] [Medline: [22861175](https://pubmed.ncbi.nlm.nih.gov/22861175/)]
2. Heron J, O'Connor T, Evans J, Golding J, Glover V, ALSPAC Study Team. The course of anxiety and depression through pregnancy and the postpartum in a community sample. *J Affect Disord* 2004 May;80(1):65-73. [doi: [10.1016/j.jad.2003.08.004](https://doi.org/10.1016/j.jad.2003.08.004)] [Medline: [15094259](https://pubmed.ncbi.nlm.nih.gov/15094259/)]
3. National IFHE. Antenatal and postnatal mental health: clinical management and service guidance. London: National Institute for Health and Care Excellence. 2014. URL: <https://www.nice.org.uk/guidance/cg192/resources/antenatal-and-postnatal-mental-health-clinical-management-and-service-guidance-pdf-35109869806789> [accessed 2022-02-26]
4. Ansariniaki M, Abounoori M, Babakhanian M. A systematic review of anxiety during pregnancy in the period of the COVID-19 pandemic. *Int J Pediatr* 2021 Feb 23;9(11):14756-14771. [doi: [10.22038/IJP.2021.54774.4321](https://doi.org/10.22038/IJP.2021.54774.4321)]
5. Moyer CA, Compton SD, Kaselitz E, Muzik M. Pregnancy-related anxiety during COVID-19: a nationwide survey of 2740 pregnant women. *Arch Womens Ment Health* 2020 Dec;23(6):757-765 [FREE Full text] [doi: [10.1007/s00737-020-01073-5](https://doi.org/10.1007/s00737-020-01073-5)] [Medline: [32989598](https://pubmed.ncbi.nlm.nih.gov/32989598/)]
6. Bayrampour H, Ali E, McNeil D, Benzies K, MacQueen G, Tough S. Pregnancy-related anxiety: a concept analysis. *Int J Nurs Stud* 2016 Mar;55:115-130. [doi: [10.1016/j.ijnurstu.2015.10.023](https://doi.org/10.1016/j.ijnurstu.2015.10.023)] [Medline: [26626973](https://pubmed.ncbi.nlm.nih.gov/26626973/)]
7. Harpel TS. Fear of the unknown: ultrasound and anxiety about fetal health. *Health (London)* 2008 Jul;12(3):295-312. [doi: [10.1177/1363459308090050](https://doi.org/10.1177/1363459308090050)] [Medline: [18579629](https://pubmed.ncbi.nlm.nih.gov/18579629/)]
8. Guardino CM, Dunkel Schetter C, Bower JE, Lu MC, Smalley SL. Randomised controlled pilot trial of mindfulness training for stress reduction during pregnancy. *Psychol Health* 2014;29(3):334-349 [FREE Full text] [doi: [10.1080/08870446.2013.852670](https://doi.org/10.1080/08870446.2013.852670)] [Medline: [24180264](https://pubmed.ncbi.nlm.nih.gov/24180264/)]
9. Wenzel A. Anxiety in Childbearing Women: Diagnosis and Treatment. Washington DC: American Psychological Association; 2011.
10. Staneva A, Bogossian F, Pritchard M, Wittkowski A. The effects of maternal depression, anxiety, and perceived stress during pregnancy on preterm birth: a systematic review. *Women Birth* 2015 Sep;28(3):179-193. [doi: [10.1016/j.wombi.2015.02.003](https://doi.org/10.1016/j.wombi.2015.02.003)] [Medline: [25765470](https://pubmed.ncbi.nlm.nih.gov/25765470/)]

11. Grigoriadis S, Graves L, Peer M, Mamisashvili L, Tomlinson G, Vigod SN, et al. A systematic review and meta-analysis of the effects of antenatal anxiety on postpartum outcomes. *Arch Womens Ment Health* 2019 Oct;22(5):543-556. [doi: [10.1007/s00737-018-0930-2](https://doi.org/10.1007/s00737-018-0930-2)] [Medline: [30523416](https://pubmed.ncbi.nlm.nih.gov/30523416/)]
12. Koelewijn JM, Sluijs AM, Vrijkotte TGM. Possible relationship between general and pregnancy-related anxiety during the first half of pregnancy and the birth process: a prospective cohort study. *BMJ Open* 2017 May 09;7(5):e013413 [FREE Full text] [doi: [10.1136/bmjopen-2016-013413](https://doi.org/10.1136/bmjopen-2016-013413)] [Medline: [28490549](https://pubmed.ncbi.nlm.nih.gov/28490549/)]
13. Preis H, Mahaffey B, Heiselman C, Lobel M. Pandemic-related pregnancy stress and anxiety among women pregnant during the coronavirus disease 2019 pandemic. *Am J Obstet Gynecol MFM* 2020 Aug;2(3):100155 [FREE Full text] [doi: [10.1016/j.ajogmf.2020.100155](https://doi.org/10.1016/j.ajogmf.2020.100155)] [Medline: [32838261](https://pubmed.ncbi.nlm.nih.gov/32838261/)]
14. Vigod SN, Dennis CL. Advances in virtual care for perinatal mental disorders. *World Psychiatry* 2020 Oct;19(3):328-329 [FREE Full text] [doi: [10.1002/wps.20775](https://doi.org/10.1002/wps.20775)] [Medline: [32931113](https://pubmed.ncbi.nlm.nih.gov/32931113/)]
15. Evans K, Rennick-Egglestone S, Cox S, Kuipers Y, Spiby H. Remotely delivered interventions to support women with symptoms of anxiety in pregnancy: mixed methods systematic review and meta-analysis. *J Med Internet Res* 2022 Feb 15;24(2):e28093 [FREE Full text] [doi: [10.2196/28093](https://doi.org/10.2196/28093)] [Medline: [35166688](https://pubmed.ncbi.nlm.nih.gov/35166688/)]
16. Hughson JP, Daly JO, Woodward-Kron R, Hajek J, Story D. The rise of pregnancy apps and the implications for culturally and linguistically diverse women: narrative review. *JMIR Mhealth Uhealth* 2018 Nov 16;6(11):e189 [FREE Full text] [doi: [10.2196/mhealth.9119](https://doi.org/10.2196/mhealth.9119)] [Medline: [30446483](https://pubmed.ncbi.nlm.nih.gov/30446483/)]
17. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. *BMC Pregnancy Childbirth* 2016 Mar 28;16:65 [FREE Full text] [doi: [10.1186/s12884-016-0856-5](https://doi.org/10.1186/s12884-016-0856-5)] [Medline: [27021727](https://pubmed.ncbi.nlm.nih.gov/27021727/)]
18. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. *J Am Med Inform Assoc* 2018 Aug 01;25(8):1089-1098 [FREE Full text] [doi: [10.1093/jamia/ocy050](https://doi.org/10.1093/jamia/ocy050)] [Medline: [29788283](https://pubmed.ncbi.nlm.nih.gov/29788283/)]
19. Bricker JB, Mull KE, Kientz JA, Vilardaga R, Mercer LD, Akioka KJ, et al. Randomized, controlled pilot trial of a smartphone app for smoking cessation using acceptance and commitment therapy. *Drug Alcohol Depend* 2014 Oct 01;143:87-94 [FREE Full text] [doi: [10.1016/j.drugalcdep.2014.07.006](https://doi.org/10.1016/j.drugalcdep.2014.07.006)] [Medline: [25085225](https://pubmed.ncbi.nlm.nih.gov/25085225/)]
20. Tripp N, Hainey K, Liu A, Poulton A, Peek M, Kim J, et al. An emerging model of maternity care: smartphone, midwife, doctor? *Women Birth* 2014 Mar;27(1):64-67. [doi: [10.1016/j.wombi.2013.11.001](https://doi.org/10.1016/j.wombi.2013.11.001)] [Medline: [24295598](https://pubmed.ncbi.nlm.nih.gov/24295598/)]
21. Google Play Store. Google Play. 2021. URL: <https://play.google.com/store> [accessed 2022-07-05]
22. App Store. Apple - App Store. 2021. URL: <https://www.apple.com/uk/ios/app-store/> [accessed 2022-07-05]
23. Powell AC, Torous J, Chan S, Raynor GS, Shwartz E, Shanahan M, et al. Interrater reliability of mHealth app rating measures: analysis of top depression and smoking cessation apps. *JMIR Mhealth Uhealth* 2016 Feb 10;4(1):e15 [FREE Full text] [doi: [10.2196/mhealth.5176](https://doi.org/10.2196/mhealth.5176)] [Medline: [26863986](https://pubmed.ncbi.nlm.nih.gov/26863986/)]
24. NHS Apps Library. NHS Apps Library. 2021. URL: <https://www.nhs.uk/apps-library/> [accessed 2022-02-22]
25. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. *Transl Behav Med* 2014 Dec;4(4):363-371 [FREE Full text] [doi: [10.1007/s13142-014-0293-9](https://doi.org/10.1007/s13142-014-0293-9)] [Medline: [25584085](https://pubmed.ncbi.nlm.nih.gov/25584085/)]
26. Brown H, Bucher T, Collins C, Rollo M. A review of pregnancy apps freely available in the Google Play Store. *Health Promot J Austr* 2020 Sep;31(3):340-342. [doi: [10.1002/hpja.270](https://doi.org/10.1002/hpja.270)] [Medline: [31225924](https://pubmed.ncbi.nlm.nih.gov/31225924/)]
27. Musgrave LM, Kizirian NV, Homer CSE, Gordon A. Mobile phone apps in Australia for improving pregnancy outcomes: systematic search on app stores. *JMIR Mhealth Uhealth* 2020 Nov 16;8(11):e22340 [FREE Full text] [doi: [10.2196/22340](https://doi.org/10.2196/22340)] [Medline: [33196454](https://pubmed.ncbi.nlm.nih.gov/33196454/)]
28. Singh A, Chander R, Mendiratta V, Singh R, Sharma A. Vitiligo and metabolic syndrome: A case control study. In: *Pigment Cell and Melanoma Research*. Hoboken, NJ: John Wiley & Sons Ltd; 2014.
29. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. *Evid Based Ment Health* 2019 Feb;22(1):4-9 [FREE Full text] [doi: [10.1136/ebmental-2018-300069](https://doi.org/10.1136/ebmental-2018-300069)] [Medline: [30635262](https://pubmed.ncbi.nlm.nih.gov/30635262/)]
30. Domnich A, Arata L, Amicizia D, Signori A, Patrick B, Stoyanov S, et al. Development and validation of the Italian version of the Mobile Application Rating Scale and its generalisability to apps targeting primary prevention. *BMC Med Inform Decis Mak* 2016 Jul 07;16:83 [FREE Full text] [doi: [10.1186/s12911-016-0323-2](https://doi.org/10.1186/s12911-016-0323-2)] [Medline: [27387434](https://pubmed.ncbi.nlm.nih.gov/27387434/)]
31. Brown H, Bucher T, Collins C, Rollo M. A review of pregnancy iPhone apps assessing their quality, inclusion of behaviour change techniques, and nutrition information. *Matern Child Nutr* 2019 Jul;15(3):e12768 [FREE Full text] [doi: [10.1111/mcn.12768](https://doi.org/10.1111/mcn.12768)] [Medline: [30569549](https://pubmed.ncbi.nlm.nih.gov/30569549/)]
32. Mind the Bump. Mind the Bump.: Smiling Mind; 2020. URL: <https://www.mindthebump.org.au/> [accessed 2022-07-06]
33. Welcome to the Pregnancy Pack. Headspace. 2021. URL: <https://www.headspace.com/articles/welcome-to-the-pregnancy-pack> [accessed 2022-07-05]
34. Wexler A, Davoudi A, Weissenbacher D, Choi R, O'Connor K, Cummings H, et al. Pregnancy and health in the age of the Internet: a content analysis of online "birth club" forums. *PLoS One* 2020;15(4):e0230947 [FREE Full text] [doi: [10.1371/journal.pone.0230947](https://doi.org/10.1371/journal.pone.0230947)] [Medline: [32287266](https://pubmed.ncbi.nlm.nih.gov/32287266/)]

35. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. *J Med Internet Res* 2018 Jun 06;20(6):e10120 [FREE Full text] [doi: [10.2196/10120](https://doi.org/10.2196/10120)] [Medline: [29875087](https://pubmed.ncbi.nlm.nih.gov/29875087/)]
36. Shen N, Levitan M, Johnson A, Bender JL, Hamilton-Page M, Jadad AAR, et al. Finding a depression app: a review and content analysis of the depression app marketplace. *JMIR Mhealth Uhealth* 2015 Feb 16;3(1):e16 [FREE Full text] [doi: [10.2196/mhealth.3713](https://doi.org/10.2196/mhealth.3713)] [Medline: [25689790](https://pubmed.ncbi.nlm.nih.gov/25689790/)]
37. Torous JB, Chan SR, Gipson SYT, Kim JW, Nguyen T, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. *Psychiatr Serv* 2018 May 01;69(5):498-500. [doi: [10.1176/appi.ps.201700423](https://doi.org/10.1176/appi.ps.201700423)] [Medline: [29446337](https://pubmed.ncbi.nlm.nih.gov/29446337/)]
38. Generalised anxiety disorder and panic disorder in adults: management. NICE: National Institute for Health and Care Excellence. London, UK; 2019 Jul 26. URL: <https://www.nice.org.uk/guidance/cg113> [accessed 2022-02-26]
39. Antenatal and postnatal mental health: clinical management and service guidance. NICE: National Institute for Health and Care Excellence. London, UK; 2020 Feb 11. URL: <https://www.nice.org.uk/guidance/cg192> [accessed 2022-02-26]
40. Marc I, Toureche N, Ernst E, Hodnett E, Blanchet C, Dodin S, et al. Mind-body interventions during pregnancy for preventing or treating women's anxiety. *Cochrane Database Syst Rev* 2011 Jul 06(7):CD007559. [doi: [10.1002/14651858.CD007559.pub2](https://doi.org/10.1002/14651858.CD007559.pub2)] [Medline: [21735413](https://pubmed.ncbi.nlm.nih.gov/21735413/)]
41. Evans K, Spiby H, Morrell JC. Non-pharmacological interventions to reduce the symptoms of mild to moderate anxiety in pregnant women. A systematic review and narrative synthesis of women's views on the acceptability of and satisfaction with interventions. *Arch Womens Ment Health* 2020 Feb;23(1):11-28 [FREE Full text] [doi: [10.1007/s00737-018-0936-9](https://doi.org/10.1007/s00737-018-0936-9)] [Medline: [30613846](https://pubmed.ncbi.nlm.nih.gov/30613846/)]
42. Van Singer M, Chatton A, Khazaal Y. Quality of smartphone apps related to panic disorder. *Front Psychiatry* 2015;6:96 [FREE Full text] [doi: [10.3389/fpsy.2015.00096](https://doi.org/10.3389/fpsy.2015.00096)] [Medline: [26236242](https://pubmed.ncbi.nlm.nih.gov/26236242/)]
43. Bol N, Helberger N, Weert J. Differences in mobile health app use: A source of new digital inequalities? *Inf Soc* 2018 Apr 26;34(3):183-193. [doi: [10.1080/01972243.2018.1438550](https://doi.org/10.1080/01972243.2018.1438550)]
44. Pregnancy Yoga App. 2021. URL: <https://www.yogibirth.com/> [accessed 2022-07-06]
45. Hypnobirthing: Pregnancy Music and Tracker. APKCombo. 2021. URL: <https://apkcombo.com/pt/hypnobirthing-pregnancy-music-tracker/com.Hypnobirth.app/>
46. Pregnancy yoga Exercises. Google Play. 2021. URL: <https://play.google.com/store/apps/details?id=com.HomeFitness.Pregnant&hl=en&gl=US> [accessed 2022-02-26]
47. Pregnancy Yoga - Prenatal Yoga. Google Play. URL: <https://play.google.com/store/apps/details?id=com.educationapps.hfpregnant&hl=en&gl=US> [accessed 2022-02-26]
48. Carry: Pregnancy Workouts. the Carry app. 2021. URL: <https://www.thecarryapp.com/> [accessed 2022-02-26]
49. Hypnobirthing: Mom Pregnancy. App Store. 2021. URL: <https://apps.apple.com/us/app/hypnobirthing-calm-birth-app/id1489680692> [accessed 2022-02-26]
50. iHypnobirth - lite. App Store. 2021. URL: <https://apps.apple.com/gb/app/ihypnobirth-lite/id452549379> [accessed 2022-02-26]
51. Music for Pregnancy Relaxation. Google Play. 2018. URL: <https://play.google.com/store/apps/details?id=com.ultimategamestudio.musicforpregnancy&hl=en&gl=US> [accessed 2022-02-26]
52. Pregnancy Music Collection. Google Play. 2021. URL: <https://play.google.com/store/apps/details?id=tw.com.tiaozhisoft.tsregnancy&hl=et> [accessed 2022-02-26]
53. Pregnancy Care Tips. Google Play. 2015. URL: <https://play.google.com/store/apps/details?id=care.pregnancytips.pregnancysymptoms&hl=en&gl=US> [accessed 2021-07-25]
54. Google Play. 2021. URL: https://play.google.com/store/apps/details?id=de.keleya.app&hl=en_GB&gl=US [accessed 2022-02-26]
55. Baby Buddy. Best Beginnings. London; 2021. URL: <https://www.bestbeginnings.org.uk/baby-buddy> [accessed 2022-07-06]
56. Kertz SJ, MacLaren Kelly J, Stevens KT, Schrock M, Danitz SB. A review of free iPhone applications designed to target anxiety and worry. *J Technol Behav Sci* 2017 Jan 11;2(2):61-70. [doi: [10.1007/s41347-016-0006-y](https://doi.org/10.1007/s41347-016-0006-y)]
57. Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M. Web-based mindfulness interventions for mental health treatment: systematic review and meta-analysis. *JMIR Ment Health* 2018 Sep 25;5(3):e10278 [FREE Full text] [doi: [10.2196/10278](https://doi.org/10.2196/10278)] [Medline: [30274963](https://pubmed.ncbi.nlm.nih.gov/30274963/)]
58. Hussain-Shamsy N, Shah A, Vigod SN, Zaheer J, Seto E. Mobile health for perinatal depression and anxiety: scoping review. *J Med Internet Res* 2020 Apr 13;22(4):e17011 [FREE Full text] [doi: [10.2196/17011](https://doi.org/10.2196/17011)] [Medline: [32281939](https://pubmed.ncbi.nlm.nih.gov/32281939/)]
59. Carlo AD, Hosseini Ghomi R, Renn BN, Areán PA. By the numbers: ratings and utilization of behavioral health mobile applications. *NPJ Digit Med* 2019;2:54 [FREE Full text] [doi: [10.1038/s41746-019-0129-6](https://doi.org/10.1038/s41746-019-0129-6)] [Medline: [31304400](https://pubmed.ncbi.nlm.nih.gov/31304400/)]
60. Evans K, Rennick-Egglestone S, Cox S, Kuipers Y, Spiby H. Remotely delivered interventions to support women with symptoms of anxiety in pregnancy: mixed methods systematic review and meta-analysis. *J Med Internet Res* 2022 Feb 15;24(2):e28093 [FREE Full text] [doi: [10.2196/28093](https://doi.org/10.2196/28093)] [Medline: [35166688](https://pubmed.ncbi.nlm.nih.gov/35166688/)]
61. Loughnan SA, Joubert AE, Grierson A, Andrews G, Newby JM. Internet-delivered psychological interventions for clinical anxiety and depression in perinatal women: a systematic review and meta-analysis. *Arch Womens Ment Health* 2019 Dec;22(6):737-750. [doi: [10.1007/s00737-019-00961-9](https://doi.org/10.1007/s00737-019-00961-9)] [Medline: [31101993](https://pubmed.ncbi.nlm.nih.gov/31101993/)]

62. Ahmadian L, Khajouei R, Kamali S, Mirzaee M. Exploring pregnant women interest and anxiety to use the Internet and their perception and trust toward online information. *Res Sq* 2019 Sep 27;1-13 [FREE Full text] [doi: [10.21203/rs.2.15291/v1](https://doi.org/10.21203/rs.2.15291/v1)]
63. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-Mental health apps in Canada: results of a modified Delphi process. *JMIR Mhealth Uhealth* 2018 Jul 09;6(7):e10016 [FREE Full text] [doi: [10.2196/10016](https://doi.org/10.2196/10016)] [Medline: [29986846](https://pubmed.ncbi.nlm.nih.gov/29986846/)]
64. Byambasuren O, Beller E, Glasziou P. Current knowledge and adoption of mobile health apps among Australian general practitioners: survey study. *JMIR Mhealth Uhealth* 2019 Jun 03;7(6):e13199 [FREE Full text] [doi: [10.2196/13199](https://doi.org/10.2196/13199)] [Medline: [31199343](https://pubmed.ncbi.nlm.nih.gov/31199343/)]
65. Digital Technology Assessment Criteria (DTAC). NHS Transformation Directorate. London, UK: NHS England URL: <https://www.nhs.uk/key-tools-and-info/digital-technology-assessment-criteria-dtac/> [accessed 2022-02-25]

Abbreviations

CBT: cognitive behavioral therapy

Edited by R Kukafka; submitted 06.07.21; peer-reviewed by K Stawarz, C Reis, UH Mohamad; comments to author 03.11.21; revised version received 11.12.21; accepted 16.12.21; published 23.03.22.

Please cite as:

Evans K, Donelan J, Rennick-Egglestone S, Cox S, Kuipers Y

Review of Mobile Apps for Women With Anxiety in Pregnancy: Maternity Care Professionals' Guide to Locating and Assessing Anxiety Apps

J Med Internet Res 2022;24(3):e31831

URL: <https://www.jmir.org/2022/3/e31831>

doi: [10.2196/31831](https://doi.org/10.2196/31831)

PMID: [35319482](https://pubmed.ncbi.nlm.nih.gov/35319482/)

©Kerry Evans, Jasper Donelan, Stefan Rennick-Egglestone, Serena Cox, Yvonne Kuipers. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 23.03.2022. 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.

Original Paper

Efficacy, Benefits, and Harms of a Self-management App in a Swedish Trauma-Exposed Community Sample (PTSD Coach): Randomized Controlled Trial

Ida Hensler¹, MSc; Josefin Sveen¹, PhD; Martin Cernvall², PhD; Filip K Arnberg¹, PhD

¹National Centre for Disaster Psychiatry, Department of Medical Sciences, Uppsala University, Uppsala, Sweden

²Psychiatry, Department of Medical Sciences, Uppsala University, Uppsala, Sweden

Corresponding Author:

Ida Hensler, MSc

National Centre for Disaster Psychiatry

Department of Medical Sciences

Uppsala University

Akademiska sjukhuset ing. 10 van. 3

Uppsala, 75185

Sweden

Phone: 46 186170896

Email: ida.hensler@neuro.uu.se

Related Article:

This is a corrected version. See correction statement: <https://www.jmir.org/2023/1/e45847>

Abstract

Background: Self-guided interventions may complement and overcome obstacles to in-person treatment options. The efficacy of app interventions targeting posttraumatic stress disorder (PTSD) is unclear, and results from previous studies on PTSD Coach—an app for managing trauma-related distress—are inconsistent.

Objective: This study investigates whether access to the Swedish version of the PTSD Coach affects posttraumatic stress, depressive, and somatic symptoms. In addition, we aim to assess the perceived helpfulness, satisfaction, negative effects, response, and remission related to PTSD Coach.

Methods: Adults who had experienced potentially traumatic events in the past 2 years were randomized (1:1) to have access to PTSD Coach (n=89) or be on the waitlist (n=90). We assessed clinical characteristics at baseline (semistructured interviews and self-rating scales) and after 3 months (self-rating scales). We analyzed the data in R software using linear mixed effects models, chi-square tests, and Fisher exact test.

Results: Intention-to-treat analyses indicated that access to PTSD Coach decreased posttraumatic stress and depressive symptoms but not somatic symptoms. More participants who had access to PTSD Coach responded with clinically significant improvement and fewer instances of probable PTSD after 3 months compared with waitlist controls. Overall, participants found that PTSD Coach was slightly to moderately helpful and moderately satisfactory. Half of the intervention group (36/71, 51%) reported at least one negative reaction related to using PTSD Coach (eg, disappointment with the app or its results, arousal of stress, or distressing memories).

Conclusions: Using PTSD Coach may trigger symptoms among a few users; however, most of them perceived PTSD Coach as helpful and satisfactory. This study showed that having access to PTSD Coach helped improve psychological trauma-related symptoms. In addition, we have discussed implications for future research and clinical practice.

Trial Registration: ClinicalTrials.gov NCT04094922; <https://clinicaltrials.gov/ct2/show/NCT04094922>

(*J Med Internet Res* 2022;24(3):e31419) doi:[10.2196/31419](https://doi.org/10.2196/31419)

KEYWORDS

PTSD; self-management app; mHealth; RCT; negative effects; mobile phone

Introduction

Background

Psychological trauma, often recognized as posttraumatic stress disorder (PTSD), is an important public health problem. Most of the world's population will be exposed to one or more traumas throughout their lifetimes [1,2]. The lifetime prevalence of PTSD is nearly 4%, and it is associated with considerable burden [2]. A substantial impairment is also noted among people who experience posttraumatic stress symptoms but do not meet the complete diagnostic criteria for PTSD [3]. Trauma survivors experience significant personal and structural barriers to seeking help, including stigma, time and resource constraints, and lack of knowledge and access to mental health care [4].

PTSD Coach is a self-management mobile app for improving knowledge of PTSD symptoms and providing coping strategies for trauma-related acute distress [5-7]. The app provides psychoeducation about the effects of trauma, a self-rating scale for posttraumatic stress, and contact information to reach professional help and support organizations. It also contains a database of self-guided exercises inspired by cognitive behavioral treatment methods such as mindfulness, stress reduction techniques, grounding, positive psychology, and cognitive restructuring [6,7]. The efficacy of app interventions that target PTSD is unclear [8]. Self-guided interventions such as PTSD Coach could not replace, but may complement, in-person treatment options [9] as stand-alone interventions or as additions to psychological or medical treatments.

PTSD Coach has shown promise as a beneficial intervention in Western countries [7]. However, the results from uncontrolled studies regarding PTSD Coach and a decrease in posttraumatic stress or depressive symptoms are inconsistent [10-12]. The results from prior randomized controlled trial (RCT) studies [6,13,14] of PTSD Coach also differ, perhaps owing to differences in the operationalization of outcomes and sample sizes. For example, using PTSD Coach with or without clinician support decreased PTSD symptoms but did not change depressive symptoms [14], and having access to PTSD Coach was related to greater improvements in PTSD symptom severity [6,13] and depressive symptoms compared with controls [6]. However, the results were inconclusive, as symptoms after the intervention did not differ compared with controls [6,13].

Users tend to like PTSD Coach. Although some disagree [13], the app was generally considered moderately to extremely helpful in US samples [5,15] and slightly to moderately helpful in a pilot study of the Swedish app [12]. Users endorse being moderately to extremely satisfied [5,15] or slightly to moderately satisfied with the app [12]. Overall satisfaction seems to be higher among smartphone owners than others, whereas perceived helpfulness do not differ [5]. Similarly, users have expressed that previous digital skills may benefit the use of PTSD Coach [16]. In addition, some users questioned whether using PTSD Coach without clinician support could be harmful [16]. Deterioration (ie, worsening of symptoms [17]) should be considered a side effect if the intervention cannot be ruled out as a probable cause [18]. To the best of our knowledge, no investigation of PTSD Coach has reported the presence or

absence of deteriorated symptoms [6,10,11,13,14]. Other possible negative effects are treatment-emergent reactions (referred to in this paper as negative reactions, ie, unwanted reactions) instigated by the use of an intervention [17]. Most researchers do not report the presence or absence of negative reactions to PTSD Coach [6,10,11,13,15]. Negative reactions such as negative emotions, psychological symptoms [16,19], and unfulfilled expectations [12] in response to content or technical issues have been reported in focus groups, interviews, and reviews on the web.

To summarize, the efficacy, benefits, and risks of self-management interventions for posttraumatic stress should be evaluated [8], especially if they are intended to be distributed without clinical support. The Swedish version of PTSD Coach has yet to be evaluated in an RCT. Sweden is a relatively sparsely populated country with high levels of smartphone use. Mental health apps may therefore be particularly well-suited for use as a complement to existing services. In addition, although exposure to potentially traumatic events and posttraumatic stress are associated with somatic symptoms and disease [20] and are subject to improvement with standard therapies such as medication or cognitive behavioral therapy [21], somatic symptoms have not been investigated concerning the use of PTSD Coach.

Study Aims

We investigated whether access to PTSD Coach affected symptoms of posttraumatic stress (primary outcome), depression, and somatic illness (secondary outcomes) in an RCT. We also conducted post hoc analyses to explore response rates, clinically significant change, and deterioration in posttraumatic stress symptoms and remission rates of probable PTSD. Finally, we investigated perceived helpfulness, satisfaction, and negative reactions associated with PTSD Coach.

Methods

Design

We conducted an RCT with a parallel group, mixed model design to compare a self-management intervention with a waitlist. The intervention group (app access) had access to PTSD Coach, and the waitlist group (waitlist) did not have access to PTSD Coach for 3 months.

Participants

Adults (aged ≥ 18 years) who resided in Sweden with Swedish verbal and written comprehension and ownership of a smartphone were eligible to be included in the study. Additional inclusion criteria were exposure to a potentially traumatic event in the past 2 years, according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria, and mild to severe posttraumatic stress symptoms (PTSD Checklist for DSM-5 total score ≥ 10). Exclusion criteria included potentially life-threatening or harmful living conditions or symptoms (eg, recurring or ongoing traumatic event exposure, severe suicidal plans or ideation, current alcohol or drug abuse, lifetime manic or hypomanic episodes, or psychotic episodes). Additional exclusion criteria were current or pending psychotherapy, medical treatment changes, and medication with

counterindications such as benzodiazepines. Minor adjustments were made after the trial commenced. Participants who screened positive for alcohol or substance abuse in early remission (<12 months) with the current treatment were accepted in the study.

The required sample size of 160 participants was determined by an a priori power analysis in G*Power (version 3.1; Heinrich-Heine-Universität Düsseldorf) [22] based on the effect size of Cohen $d=0.5$ in the pilot study [12] and anticipated attrition of up to 25%.

Procedures

Enrollment began in May 2019 and ended in June 2020. We collected data nationwide in Sweden from Uppsala University. Study data and email invitations were managed using REDCap (Research Electronic Data Capture; [23]) hosted at Uppsala University.

Potential participants were recruited through social media advertisements linked to a web-based screening questionnaire. Eligible participants provided an email, received the consent form, and provided written informed consent. We informed participants how their data were managed and that participation was voluntary and confidential. Consenting participants booked an appointment for a phone interview with a member of the research team to confirm eligibility and assess psychiatric symptoms.

Subsequently, participants completed a baseline questionnaire before randomization. Participants randomized to access to the app were emailed written instructions for downloading PTSD Coach and instructed to use the app as they pleased. Participants who requested further guidance were encouraged to explore the app to identify helpful content. Participants on the waitlist received written notice through email that they would gain access to the app after the first follow-up assessment.

We called all participants and offered the opportunity to ask questions regarding the study or technical support, 7 days after randomization. All participants responded to daily assessments for 21 days in a separate investigation [24] and received a follow-up questionnaire with the primary and secondary outcomes 3 months later. Participants with access to the app responded to additional questions regarding helpfulness, satisfaction, and negative reactions. Participants were compensated with gift cards to the cinema after completing the follow-up questionnaire. Participants who left the study or were excluded before the follow-up assessment received a gift card.

Materials

The Intervention

The mobile app PTSD Coach was developed by the Veterans' Affairs National Center for PTSD and the Department of Defense's DHA Connected Health [5,25]. The resources in PTSD Coach are divided into four sections: *learn* (psychoeducation about posttraumatic stress, treatment, and coping in families), *track* (symptom self-evaluation with rating history and automatic feedback), *manage symptoms* (exercises for distress management inspired by cognitive behavioral therapy), and *get support* (contact information for crisis resources, professional assessment and treatment, and platforms

and advice promoting social support). Self-guided exercises in the app are prompted by voice, video, or text.

The Swedish version [12] was translated and adapted from the American original to a Swedish civilian context [5,6,13]. A team of 6 clinical psychologists, researchers, and psychology students conducted the first translation for the pilot trial of the app [12], and the authors revised the adaptation based on a version update of the original and input from Swedish users. During this trial, the Swedish version of PTSD Coach was not publicly available on app stores; therefore, access to the app could be restricted to participants with access to the app. However, the American version of PTSD Coach was available on app stores.

Posttraumatic Stress

The primary outcome, posttraumatic stress, was assessed at screening, baseline, and follow-up with the Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (PCL-5) [26] in Swedish [27]. The PCL-5 is a 20-item self-report measure that maps directly onto the DSM-5 PTSD symptoms. The items are rated on a 5-point scale ranging from 0 (*not at all*) to 4 (*extremely*) [26], yielding a total sum of 0 to 80. The Swedish version was developed using a standard back-translation process [28] and corresponds well to the gold standard assessment of PTSD symptoms [27]. A total sum of ≥ 31 to 33 points may indicate probable PTSD in both the original and Swedish versions [27,29]. A ≥ 10 -point difference in the total sum score on previous versions of the PTSD Checklist has been estimated to equate with clinically significant change [6,13,14,30], and the range for clinically significant change is presumably similar for PCL-5 [29].

Depressive Symptoms

The secondary outcome depressive symptoms was assessed at baseline and follow-up with the Swedish version [31] of the Patient Health Questionnaire (PHQ)-9 [32]. The questionnaire is a widely used 9-item self-report measure of the DSM-5 criteria for depression with an additional item assessing functional impairment. The suggested cut-offs for probable depression range from 9 to 12 points [32,33]. The Swedish version has demonstrated satisfactory psychometric properties [31].

Somatic Symptoms

The secondary outcome, somatic symptoms, was assessed at baseline and follow-up with the Swedish version [34] of the PHQ-15 [35]. The self-rating scale consists of 15 items that measure the most common somatic symptoms reported in primary health care. The Swedish version of the PHQ-15 has favorable psychometric properties for the quantification of somatization in Swedish and similar populations, with suggested cut-off scores for symptom severity at 0 to 4 indicating minimal, 5 to 9 indicating low, 10 to 14 indicating moderate, and ≥ 15 indicating high symptom severity [34].

Helpfulness, Satisfaction, and Negative Effects

Perceived helpfulness and overall satisfaction with PTSD Coach were measured with a Swedish version [12] of the PTSD Coach

survey [5] at follow-up. The survey includes 14 questions that assess the perceived helpfulness of the app and 1 item asking about user satisfaction. [Multimedia Appendix 1](#) provides all items. Negative reactions were measured using the Negative Effects Questionnaire [36]. The original version was developed in Sweden and is available in several languages to assess the negative effects of psychological interventions. In this study, references to therapists and treatment in the items were changed to the app. Respondents rated whether they had experienced 20 potential negative reactions during the period they had access to PTSD Coach, how negative reactions affected them, and the probable cause of the reaction (ie, “other circumstances” or “used the app”). [Multimedia Appendix 2](#) shows all items.

Demographic and Clinical Characteristics

During the phone interview, the interviewer explored exposure to potentially traumatic events with open-ended questions (eg, “Please briefly tell me about the kind of severe event that you experienced at that time” and “What happened? You don’t need to tell me everything; I just need to understand what kind of event it was”) and categorized answers into categories from the Life Events Checklist for DSM-5 [37] in Swedish [27]. We assessed psychiatric comorbidity during the baseline phone interview with the Mini International Neuropsychiatric Interview [38] (Swedish version 7.0.0). We asked questions regarding demographic information, previous smartphone use, and previous treatment in the baseline questionnaire. Furthermore, participants provided information regarding treatment changes, the use of PTSD Coach, or other self-management apps in the follow-up questionnaire.

Randomization

An external statistician generated the allocation sequence in R software (R Foundation for Statistical Computing; [Multimedia Appendix 3](#)): a random number table with equal allocation to access PTSD Coach or waitlist (1:1) with an unstratified block design fixed at 20 allocations. The first author (IH) uploaded the random number table without reviewing it into the Uppsala University REDCap randomization tool before data collection. The block size, R-script, and allocation sequence were concealed to members of the research team who enrolled participants (the first author, a clinical psychologist, and 2 psychology students) or randomized participants (the first author, a PhD student, and a clinical psychologist). The condition was revealed in the REDCap graphical user interface, nonblinded but unalterable, after randomization.

Data Analysis

We analyzed data in R (versions 3.5.1 and 4.0.5) using linear mixed effects models (*nlme* package v3.1-141). We specified 3 separate multiple regression models a priori with the direct and interaction effects of condition and time on posttraumatic stress, depressive symptoms, and somatic symptoms. All randomized participants were included in the intention-to-treat

and per-protocol analyses. We replaced single missing item ratings in the outcomes with the individual’s mean item rating ($n=24$ had 1 missing item for the PHQ-15 at baseline, and $n=1$ had 1 missing value for the PHQ-15 post intervention). Missing data for the outcomes at follow-up were assumed to be missing at random conditional on the baseline PCL-5 scores, as dropout was related to the baseline PCL-5 scores but not to any other baseline variables. We addressed the missing data using multiple imputations in the primary analyses, and we conducted sensitivity analyses, as reported in [Multimedia Appendix 4](#). We included baseline posttraumatic stress as a predictor and imputed each outcome in 500 data sets (10 iterations) with predictive mean matching [39] (*mice* package v3.13.0 and *miceadds* package v3.11-6). We report pooled parameter estimates across all imputations. We calculated model-based, between-group standardized mean differences (Cohen *d*) with 95% CIs on the pooled within-group SD at baseline [40]. Negative reactions that were caused by circumstances other than using PTSD Coach were recorded as 0 (absent).

We conducted post hoc tests to assess whether the number of participants who reported clinically significant improvement or deterioration in posttraumatic stress (a 10-point difference on the PCL-5 from baseline to follow-up) or screened positive for PTSD (≥ 31 points on the PCL-5) at baseline and follow-up differed between conditions using chi-square tests. In addition, we explored whether remission from probable PTSD (≥ 31 to < 31 points) or development of probable PTSD (from < 31 to ≥ 31 points) differed between conditions using chi-square and Fisher exact tests.

Ethics Approval

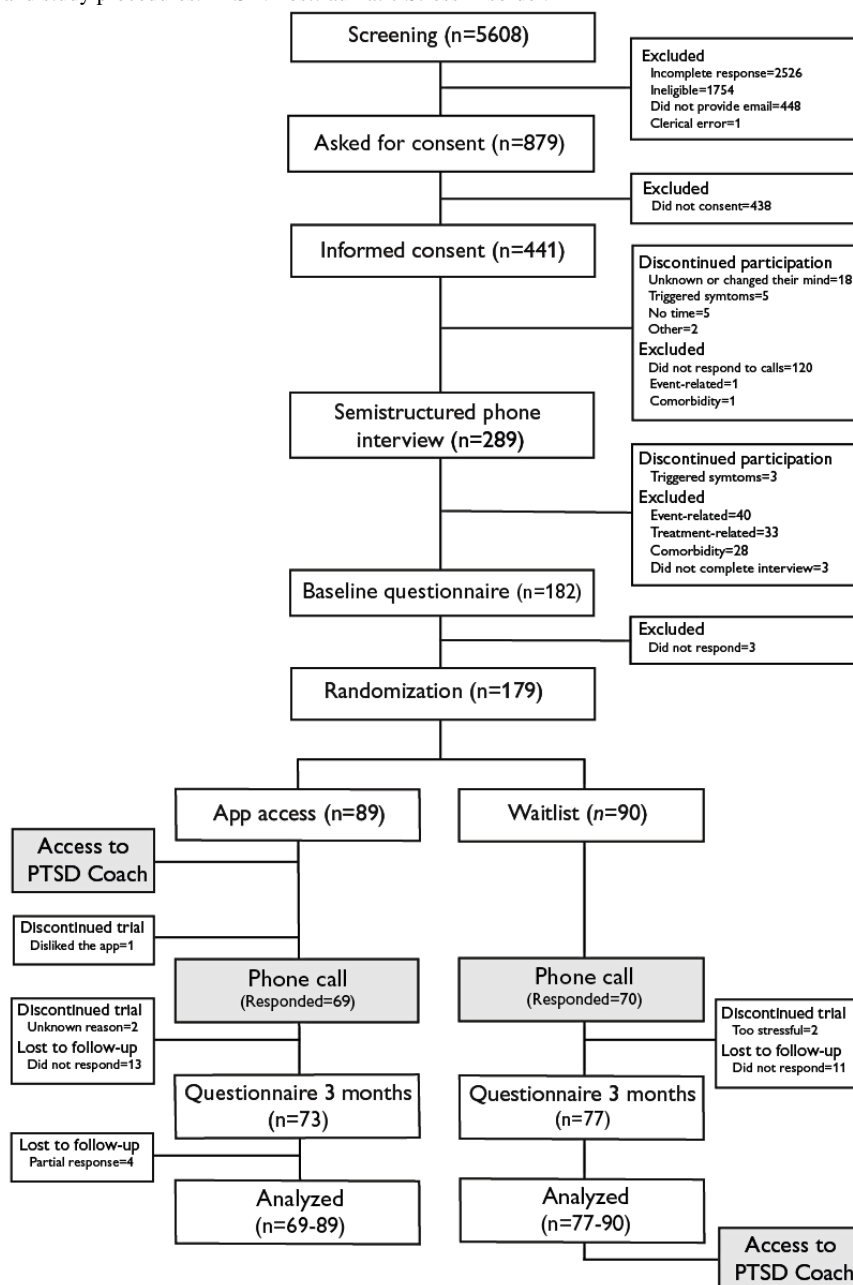
The regional ethical review board in Uppsala, Sweden, approved the study procedures before data collection (reference 2018/319).

Results

Overview

A total of 179 adults were included in the study, with 5 participants discontinuing participation after randomization ([Figure 1](#)). Using Welch *t* test, we detected no difference in posttraumatic stress between participants who discontinued participation ($n=5$; mean 50.60, SD 18.82) and those who completed the follow-up ($n=150$; mean 35.96, SD 15.90; $t_{4,2}=1.72$; $P=.16$). Participants who discontinued participation did not differ in baseline depressive symptoms ($t_{4,1}=0.63$; $P=.56$) or somatic symptoms ($t_{4,3}=0.35$; $P=.75$) compared with completers. However, baseline posttraumatic stress was higher among participants who were lost to follow-up ($n=24$; mean 42.96, SD 13.54) compared with completers ($t_{34}=2.29$; $P=.03$). Participants who were lost to follow-up did not differ in baseline depressive ($t_{32}=0.05$; $P=.96$) or somatic symptoms ($t_{30}=0.71$; $P=.48$) compared with completers.

Figure 1. Participant flow and study procedures. PTSD: Posttraumatic Stress Disorder.



Sample Characteristics

Most participants were women who used their smartphones daily (Table 1). The participants were on average 42.78 years (SD 10.90 years; app access mean 43.42, SD 10.60; waitlist mean 42.15, SD 11.21). We assessed the most severe potentially traumatic event in the past 2 years as perceived by the participants. Positive screening for disorders was assessed using the Mini International Neuropsychiatric Interview (Swedish version 7.0.0).

Most participants had experienced a potentially traumatic event firsthand in the past 2 years (Table 1), with 38 participants (app access: n=21; waitlist: n=17) reporting that their lifetime’s worst potentially traumatic event occurred more than 2 years ago. Several participants (app access=38; waitlist=39) had received previous psychological treatment or counseling after exposure to a potentially traumatic event. The levels of posttraumatic stress, depressive, and somatic symptoms were moderate to high (Table 2), and probable psychiatric disorders were common (Table 1).

Table 1. Demographic characteristics and clinician assessment of potentially traumatic events and current psychiatric conditions (N=179).

Characteristic	Total, n (%)	App access (n=89), n (%)	Waitlist (n=90), n (%)
Gender			
Women	164 (91.6)	80 (89.9)	84 (93.3)
Other ^a	15 (8.4)	9 (10.1)	6 (6.7)
Civil status			
Married or cohabitating	87 (48.6)	43 (48.3)	44 (48.9)
Single	74 (41.3)	39 (43.8)	35 (38.9)
Other ^b	18 (10.1)	7 (7.9)	11 (12.2)
Completed education			
University degree	100 (55.9)	54 (60.7)	46 (51.1)
Senior high school diploma (gymnasium)	59 (33)	26 (29.2)	33 (36.7)
Other ^c	20 (11.2)	8 (9)	11 (12.2)
Occupation			
Employed full-time or part-time	115 (64.3)	59 (66.3)	56 (62.2)
Sick leave or unemployed	32 (17.9)	15 (16.9)	17 (18.9)
Student	18 (10.1)	9 (10.1)	9 (10)
Other (eg, retired)	14 (7.8)	6 (6.7)	8 (8.9)
Daily smartphone use			
>2 hours	146 (81.6)	76 (85.4)	70 (77.8)
≤2 hours	33 (18.4)	13 (14.6)	20 (22.2)
Less than daily	0 (0)	0 (0)	0 (0)
Exposure proximity			
Experienced	99 (55.3)	48 (53.9)	51 (56.7)
Witnessed	49 (27.4)	21 (23.6)	28 (31.1)
Was told or repeated exposure	31 (17.3)	20 (22.5)	11 (12.2)
Event category			
Sudden, violent, or unexpected death	44 (24.6)	23 (25.8)	21 (23.3)
Physical assault or violence	35 (19.5)	17 (19.1)	18 (20)
Sexual assault or violence	32 (17.9)	14 (15.7)	18 (20)
Life-threatening illness or injury	26 (14.5)	9 (10.1)	17 (18.9)
Accident (vehicle or other)	20 (11.2)	14 (15.7)	6 (6.7)
Other stressful events ^d	22 (12.3)	12 (13.5)	10 (11.1)
Disorder			
PTSD^e	99 (55.3)	51 (57.3)	48 (53.3)
Subtype depersonalization or derealization	22 (12.3)	11 (12.4)	11 (12.2)
Subtype delayed onset	12 (6.7)	5 (5.6)	7 (7.8)
Suicidality (past month)^f	77 (43)	36 (40.4)	41 (45.6)
Lifetime suicide attempt	43 (24)	23 (25.8)	20 (22.2)
Depressive episode (current)	57 (31.8)	27 (30.3)	30 (33.3)
Anxiety disorder ^g	65 (36.3)	32 (36)	33 (36.7)
Other condition ^h	18 (10.1)	12 (13.5)	6 (6.7)

^aMen, other, or preferred not to answer.

^bIn a relationship (without cohabitation) or cohabitating with parents or other adults.

^cIncomplete junior or senior high school diploma or complete vocational degree.

^dNatural disasters, exposure to explosions, fires, dangerous chemicals, war zones or combat, captivity, or other severe experience.

^ePTSD: posttraumatic stress disorder.

^f>0 points on the Mini International Neuropsychiatric Interview suicidality scale, module B.

^gPanic disorder, agoraphobia, social anxiety disorder, or generalized anxiety disorder.

^hBulimia nervosa, binge-eating disorder, obsessive-compulsive disorder, or substance or alcohol abuse (past 12 months).

Table 2. Clinical characteristics: observed self-rating of symptoms at baseline and after the intervention (N=179; app access: n=89; waitlist: n=90).

Symptoms and condition	Baseline, mean (SD)	After the intervention, mean (SD) ^a
Posttraumatic stress (PCL-5^b)		
All	37.31 (15.94)	32.33 (18.44)
App access	36.44 (16.49)	27.47 (17.61)
Waitlist	38.17 (15.42)	36.95 (18.13)
Depressive symptoms (PHQ-9^c)		
All	10.88 (6.68)	10.02 (6.90)
App access	10.65 (6.79)	8.60 (6.07)
Waitlist	11.11 (6.59)	11.36 (7.40)
Somatic symptoms (PHQ-15^d)		
All	12.10 (5.56)	11.49 (5.47)
App access	11.43 (5.83)	10.48 (5.61)
Waitlist	12.77 (5.22)	12.44 (5.19)

^aAttrition at follow-up: n=29; app access: n=16; waitlist: n=13.

^bPCL-5: Posttraumatic Symptom Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

^cPHQ-9: Patient Health Questionnaire-9, depression.

^dPHQ-15: Patient Health Questionnaire-15, somatic symptoms.

Confounding Factors

There was some contamination in both groups during the trial. At follow-up, 4 participants on the waitlist reported having used PTSD Coach (presumably the English version of the app), and 7 participants (app access=4; waitlist=3) reported using a self-management app other than PTSD Coach. We also detected potentially confounding factors—participants started psychological treatment (app access=10; waitlist=10), changed their medication (app access=8; waitlist=10), or started a new medication (app access=10; waitlist=8). Moreover, 26 people sought professional help, such as medical or psychological treatment, related to their trauma (app access=17; waitlist=9) during the intervention period. At follow-up, 17 participants with access to the app stated that they had not used PTSD Coach.

Primary Outcome: Posttraumatic Stress

Access to PTSD Coach led to a greater decrease in posttraumatic stress after 3 months compared with the waitlist (Figure 2 and Table 3). The standardized mean difference was small (Cohen $d=-0.45$, 95% CI -0.70 to -0.20). The results from the sensitivity analyses per the protocol and excluding contamination (Multimedia Appendix 4) yielded highly similar results.

Furthermore, we explored the rates of improvement, response, and remission (Table 4). Participants with access to the app were more likely to show clinically significant improvement ($\chi^2_{1,150}=4.62$; $P=.03$) and less likely to fulfill the criteria for probable PTSD than participants on the waitlist after 3 months ($\chi^2_{1,150}=7.74$; $P=.005$). However, we detected no difference between conditions in remission from probable PTSD (Table 4).

Figure 2. Posttraumatic stress from baseline to follow-up. The panels present pooled, model-based group means and 95% CIs and the distributions of unimputed observations (N=179). PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

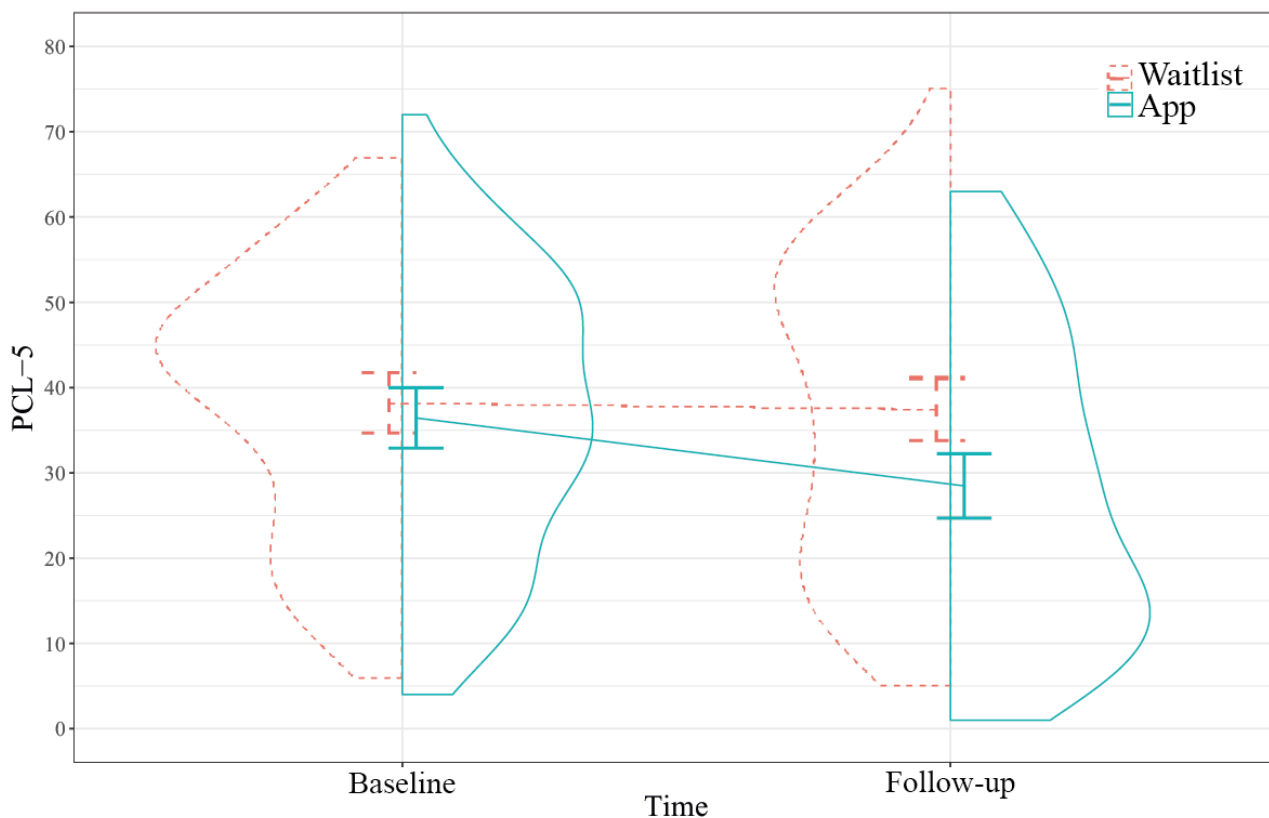


Table 3. Parameter estimates, SEs, and CIs for the multiple regression analyses of symptoms, condition, and time (pooled data). We imputed missing data at follow-up (n=29) using predictive mean matching (500 data sets; 10 iterations; N=179).

Outcome and effect	B (SE; 95% CI)	P value
Posttraumatic stress		
Intercept	38.17 (1.78; 34.66 to 41.67)	<.001
Time	-0.74 (1.55; -3.79 to 2.31)	.63
Condition ^a	-1.73 (2.53; -6.70 to 3.25)	.49
Condition×time ^b	-7.23 (2.22; -11.60 to -2.85)	.001
Depressive symptoms		
Intercept	11.11 (0.71; 9.71 to 12.51)	<.001
Time	0.33 (0.71; -1.07 to 1.72)	.65
Condition ^a	-0.46 (1.01; -2.44 to 1.52)	.65
Condition×time ^b	-2.34 (1.01; -4.32 to -0.35)	.02
Somatic symptoms		
Intercept	12.77 (0.58; 11.63 to 13.90)	<.001
Time	-0.34 (0.54; -1.40 to 0.72)	.53
Condition ^a	-1.33 (0.82; -2.94 to 0.27)	.10
Condition×time ^b	-0.72 (0.76; -2.22 to 0.79)	.35

^a0=waitlist, 1=access to the PTSD Coach.

^bFrom baseline to follow-up after 3 months.

Table 4. Access to PTSD Coach, remission, deterioration, and improvement of posttraumatic stress. App access participants had access to PTSD Coach for 3 months, whereas waitlist participants did not (n=150-179; app access: n=73-89; waitlist: n=77-90).

Outcome	App access		Waitlist		P value (χ^2 ^a)
	Value, n (%)	Value, N	Value, n (%)	Value, N	
Probable PTSD ^b at baseline ^c	57 (64)	89	63 (70)	90	.49
Probable PTSD at follow-up ^b	27 (37)	73	47 (61)	77	.005
Remission from PTSD ^d	18 (25)	73	12 (16)	77	.21
Development of PTSD ^e	3 (4)	73	6 (8)	77	.33
Clinically significant improvement ^f	30 (41)	73	18 (23)	77	.03
Clinically significant deterioration ^g	9 (12)	73	20 (26)	77	.06

^aDevelopment of posttraumatic stress disorder was evaluated with the Fisher exact test.

^bPTSD: posttraumatic stress disorder.

^c≥31 points on the Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.

^dTransitioned from ≥31 to <31 points on PCL-5 from 0 to 3 months.

^eTransitioned from <31 to ≥31 points on PCL-5 from 0 to 3 months.

^f≥10 point decrease on Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition from 0 to 3 months.

^g≥10 point increase on the Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition from 0 to 3 months.

Secondary Outcomes

Depressive and Somatic Symptoms

Access to PTSD Coach conferred a greater decrease in depressive symptoms after 3 months compared with the waitlist (Figure 3 and Table 3). The standardized mean difference was

small (Cohen $d=-0.35$, 95% CI -0.62 to -0.07). The sensitivity analyses per protocol and excluding contamination for depressive symptoms yielded highly similar results (Multimedia Appendix 4). We detected no difference in somatic symptoms based on access to PTSD Coach during 3 months (Figure 4 and Table 3). The standardized mean difference was trivial (Cohen $d=-0.13$, 95% CI -0.38 to 0.12).

Figure 3. Depressive symptoms from baseline to follow-up. The panels present pooled, model-based group means and 95% CIs and the distributions of unimputed observations (N=179). PHQ-9: Patient Health Questionnaire-9.

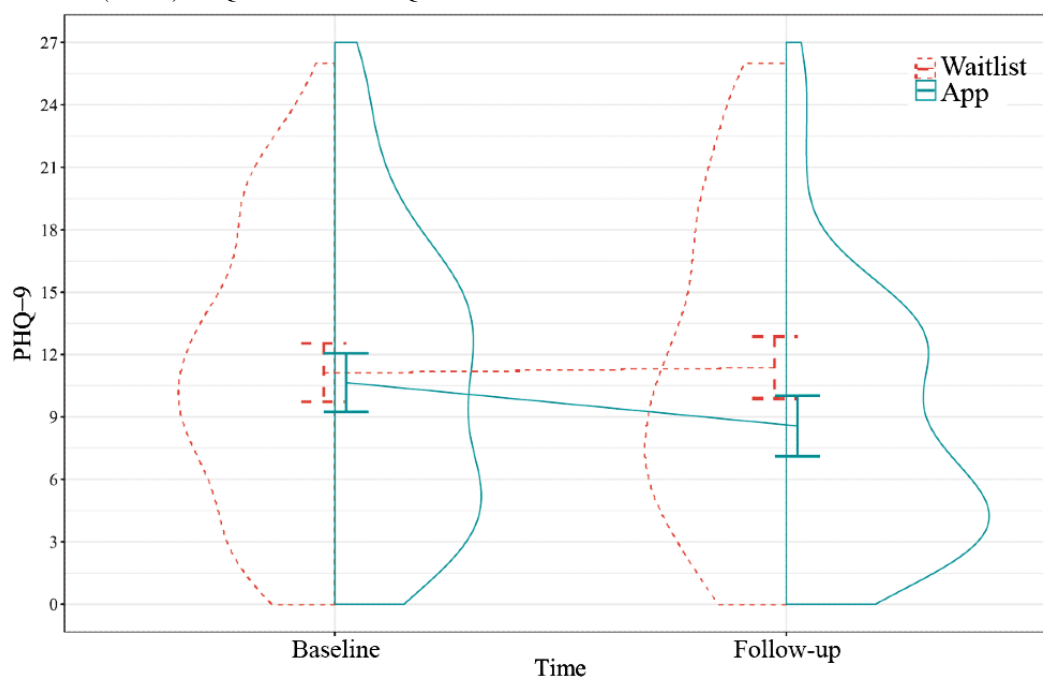
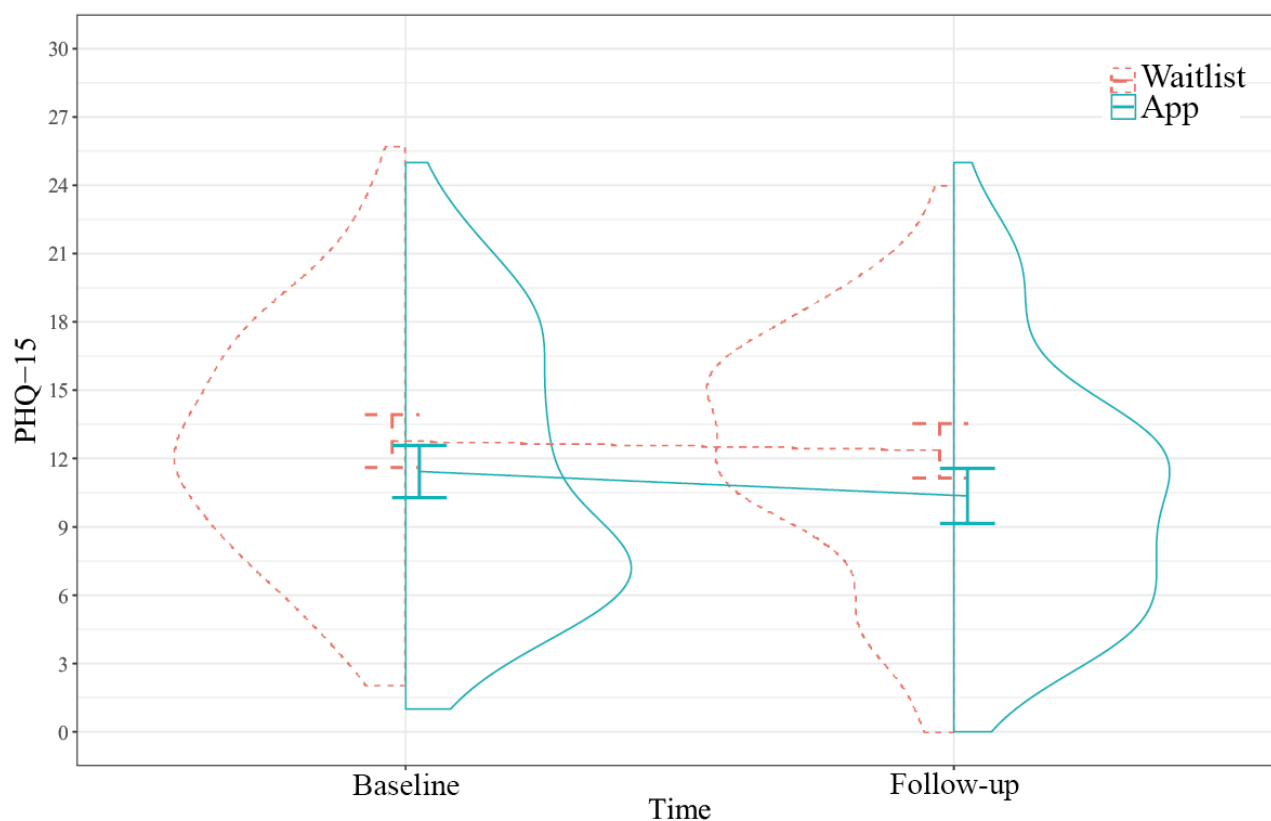


Figure 4. Somatic symptoms from baseline to follow-up. The panels present pooled, model-based group means and 95% CIs and the distributions of unimputed observations (N=179). PHQ-15: Patient Health Questionnaire-15.



Helpfulness and Satisfaction

Among participants with access to the app, 4 did not complete the PTSD Coach survey. Average ratings on helpfulness items ranged from 1.29 to 2.03 (Multimedia Appendix 1), which indicates that participants with access to PTSD Coach found the app slightly to moderately helpful. The helpful aspects of PTSD Coach are presented in Multimedia Appendix 5. The average sum score on helpfulness was 23.11 (SD 14.32; n=71). Most participants (50/69, 72%) were moderately or very satisfied with the app (n=69, mean 2.22, SD 1.07).

Negative Effects

Only 2 participants with access to the app did not respond to the Negative Effects Questionnaire, whereas 35 reported no negative reactions. In total, 36 people reported 1 to 8 negative reactions that, on average, affected them moderately (mean 2.07, SD 0.86), with an average total sum of 7.44 (SD 6.91; n=36). The most common negative reactions with moderate to extreme impact (Multimedia Appendix 6) were related to the design and evaluation of PTSD Coach, such as unfulfilled expectations on the app (14/71, 20%), its results (10/71, 14%), and perceiving the app as unmotivating (13/71, 18%) or confusing (8/71, 11%). Up to 13% (9/71) of participants experienced negative reactions in the form of psychological symptoms (Multimedia Appendix 2). Moderate to extreme symptom-related reactions included increased stress (7/71, 10%), arousal of distressing memories (6/71, 8%), anxiety (5/71, 7%), and symptom deterioration (5/71, 7%). In contrast, no one reported increased suicidality associated with the PTSD Coach or dependency on the app (Multimedia Appendix 2). Moreover,

we detected no difference in symptom deterioration between participants with access to the app and those on the waitlist (Table 4).

Discussion

Principal Findings

We conclude that access to PTSD Coach during 3 months decreased posttraumatic stress and depressive symptoms but not somatic symptoms, as compared with a waitlist control. Users perceived PTSD Coach as slightly to moderately helpful and moderately satisfactory. We found no evidence of symptom deterioration among users of PTSD Coach compared with the waitlist, and the most commonly reported negative reactions were related to the evaluation of the app and its design.

The participants' severity of posttraumatic stress was comparable [10,12] or lower [6,13,14] than the symptom burden in previous studies. Nonetheless, the treatment effect for posttraumatic stress was comparable with the response (range 5-11 points) observed in most previous studies [6,12-14]. Depressive symptoms decreased more than [10,14] or similarly [6,12] to the response in previous studies. Although participants exhibited moderate somatic symptoms, the app primarily targeted psychological distress, which could explain the lack of significant results. In addition, the origin of participants' mental health and somatic symptoms may be disparate.

The perceived helpfulness and satisfaction were generally lower than the American version of PTSD Coach [5] and greater than in the Swedish pilot study [12]. We speculate that the

participants' lower ratings of helpfulness compared with their satisfaction ratings may, in part, reflect modest expectations of the potential benefits of a smartphone app. Nonetheless, resolving technical issues and further cultural adaptation of the content may improve perceived helpfulness. Similar to participants in clinical trials of psychological treatments [36], approximately half of the intervention participants reported at least one negative reaction associated with using PTSD Coach. Using PTSD Coach may have triggered trauma-related symptoms (eg, stress, anxiety, and distressing memories), which would counteract the app's purpose. However, 87% (62/71) did not experience symptom-related negative reactions related to app use. Furthermore, we detected no differences in self-rated deterioration (clinically significant deterioration or development of probable PTSD) in posttraumatic stress between participants with and without access to PTSD Coach. Instead, access to the PTSD Coach was associated with self-rated response and recovery (ie, more instances of clinically significant improvement and fewer cases of probable PTSD) from posttraumatic stress compared with the waitlist.

Strengths and Limitations

Our rigorous evaluation of PTSD Coach, an app based on accumulated clinical expertise and research, illuminates the potential positive and negative effects of digital health interventions for posttraumatic stress. The strength of the procedure is that participants on the waitlist and with access to the app received equal attention from the research team. We opted for an inactive waitlist under genuine equipoise as to whether PTSD Coach would be superior, and the results would also, although imperfectly, represent the impact of access to PTSD Coach compared with situations in which professional care may be temporarily unavailable to increase ecological validity; for example, during waitlist for psychological treatment or in the aftermath of mass disaster situations. We did not specifically restrict participants from using sources of support, only other psychological treatments, which would reflect a situation in which access to expert treatment is scarce, but people use other sources of support. Nevertheless, the study design does not permit distinguishing whether PTSD Coach might function as a placebo, which has been discussed as a risk, particularly in smartphone apps [41], as it would entail using a sham or genuinely inactive app as a comparator. In addition, some intervention participants stated that they never used PTSD Coach, which may reflect that they did not receive the intervention. However, some of these participants reported that they had used the app when asked during the first week of the intervention period. To better understand symptom improvement in participants associated with app use, objective app data would be beneficial. We could have ascertained that the users had received the intervention and increased opportunities for gaining app access by sending instructions through multiple channels (eg, email, letter, and text).

We hope that the results of the intervention and negative effects can aid clinicians and users in making balanced and informed decisions about using self-management apps. We know that negative effects occurred, but not to what extent they persisted. Considering that the people lost to follow-up had elevated initial symptoms of posttraumatic stress and depression, negative

reactions might be underrepresented. Assessment of negative reactions in the waitlist condition would have enabled a controlled comparison. Nevertheless, the questionnaires would be dissimilar by necessity, as participants on the waitlist did not have access to PTSD Coach.

Another strength is that we used multiple methods of assessment at baseline and assessed both positive and negative outcomes with psychometrically sound instruments. However, the lack of specificity of the symptom measurement could have introduced bias to the promising results: self-ratings on the PCL-5 may not discriminate between posttraumatic stress and depressive symptoms [26]. The PCL-5 has greater sensitivity than specificity for detecting PTSD than the gold standard clinician assessment [26]. Furthermore, the assessment of outcomes is at risk of common method bias; for instance, the clinician screening for PTSD (n=99) differed from self-rated probable PTSD (n=120) at baseline. Multiple forms of assessment (clinician-assessed and self-rated) at follow-up would have improved the validity of the results.

Finally, retention was high, and we did not find evidence of bias associated with missing outcome data. We acknowledge that the predominantly female sample limits the generalizability of the results. Still, the sample was comparable with national estimates regarding marriage or cohabitation (59%) [42], university education (38%-50%) [43], and employment or studies (65%-70%) [44]. We conducted the study in an industrialized, Western society with high rates of smartphone ownership and tax-funded, low-cost mental health care. Similar interventions may need adaptations to promote symptom reduction among other genders or societal contexts.

Research Implications

In previous studies, the subjectively reported use of PTSD Coach was unrelated to changes in outcomes [6,13]. Therefore, access to PTSD Coach may affect symptoms by moderating processes other than usage frequency. The speculated mechanisms of change by using PTSD Coach include psychoeducation, coping skills, symptom awareness, and social support, which may assist treatment-seeking [6,7,10]. The content of PTSD Coach prompts seeking qualified care, which, if provided, would benefit symptom alleviation. We have limited information regarding adherence to the intervention or use of PTSD Coach during the trial, and future studies would benefit from recording objective use data or contextual information, such as when and where the use of a self-management app successfully mitigates short-term distress. Future research into the mechanisms of change and moderating processes would greatly advance the field and future design of effective mobile self-management interventions for populations in need. We also encourage others to explore the extent and persistence of negative effects concerning psychological or self-management interventions.

Clinical Implications

Given the observed efficacy, benefits, and harms of PTSD Coach, an unguided self-management app has limited potential to cure PTSD and is unlikely to replace evidence-based psychological treatments for posttraumatic stress when such treatment is warranted. Nevertheless, gaining access to PTSD

Coach seems superior to no intervention. Therefore, we believe the results support the feasibility of distributing PTSD Coach, along with other means of support, particularly after events that may temporarily restrict access to support or overwhelm treatment resources, such as pandemics, mass casualty situations, or other disasters. We speculate that access to similar content could boost the skills and knowledge acquired in previous counseling or treatment.

The impact of access to PTSD Coach during 3 months resulted in a slight reduction of posttraumatic stress. Although the average symptom reduction did not indicate a clinically significant change in posttraumatic stress, clinically significant improvement was more common among participants with access

to PTSD Coach. We believe that the benefits of disseminating a free, low-intensity intervention for potentially reducing trauma-related distress outweigh the risk of the negative effects we recorded. Nonetheless, we advise that users are informed about the possibility of negative effects to make an informed choice before using apps for mental health. Moreover, to adjust expectations, clinicians could inform users of the extent to which PTSD Coach may be beneficial or helpful. PTSD Coach is a low-risk, helpful, and effective intervention for reducing posttraumatic stress and depressive symptoms. Access to PTSD Coach may complement other psychological, medical, and social interventions for PTSD or provide an attainable first step for survivors of psychological trauma to learn, cope, and begin their road to recovery.

Acknowledgments

The authors thank the participants who offered their time and made an effort to participate in the study. The authors also thank Tove Ringmar for assisting with data collection and Hans Arinell for the independent statistical replication of the main findings. This study was funded by grants from the Swedish Board of Health and Welfare and the Swedish Crime Victim Authority (grant 03448/2017). Data are available upon request [45].

Conflicts of Interest

None declared.

Multimedia Appendix 1

Descriptive statistics of the PTSD Coach survey.

[DOCX File, 23 KB - [jmir_v24i3e31419_app1.docx](#)]

Multimedia Appendix 2

Descriptive statistics of the Negative Effects Questionnaire (PTSD Coach version).

[DOCX File, 25 KB - [jmir_v24i3e31419_app2.docx](#)]

Multimedia Appendix 3

Randomization code in R software.

[DOCX File, 21 KB - [jmir_v24i3e31419_app3.docx](#)]

Multimedia Appendix 4

Parameter estimates, standard errors (SEs), and confidence intervals (CIs) for the multiple regression analysis of symptoms by condition and time (sensitivity analysis).

[DOCX File, 26 KB - [jmir_v24i3e31419_app4.docx](#)]

Multimedia Appendix 5

Percentages of moderate to extremely helpful aspects of the PTSD Coach (n=71). Helpfulness was assessed after 3 months of access to the PTSD Coach. Helpfulness was rated as 0 (not at all), 1 (slightly), 2 (moderately), 3 (very), and 4 (extremely).

[DOCX File, 83 KB - [jmir_v24i3e31419_app5.docx](#)]

Multimedia Appendix 6

Percentages of moderate to extreme negative reactions to the PTSD Coach (n=71). Negative reactions were assessed after 3 months of access to PTSD Coach. The impact of negative reactions was rated as 0 (not at all), 1 (slightly), 2 (moderately), 3 (very), and 4 (extremely). NEQ: Negative Effects Questionnaire.

[DOCX File, 83 KB - [jmir_v24i3e31419_app6.docx](#)]

Multimedia Appendix 7

Consort E-Health Checklist (V.1.6.1).

[PDF File (Adobe PDF File), 1203 KB - [jmir_v24i3e31419_app7.pdf](#)]

References

1. Frans O, Rimmö PA, Aberg L, Fredrikson M. Trauma exposure and post-traumatic stress disorder in the general population. *Acta Psychiatr Scand* 2005 Apr;111(4):291-299. [doi: [10.1111/j.1600-0447.2004.00463.x](https://doi.org/10.1111/j.1600-0447.2004.00463.x)] [Medline: [15740465](https://pubmed.ncbi.nlm.nih.gov/15740465/)]
2. Kessler RC, Aguilar-Gaxiola S, Alonso J, Benjet C, Bromet EJ, Cardoso G, et al. Trauma and PTSD in the WHO world mental health surveys. *Eur J Psychotraumatol* 2017;8(sup5):1353383 [FREE Full text] [doi: [10.1080/20008198.2017.1353383](https://doi.org/10.1080/20008198.2017.1353383)] [Medline: [29075426](https://pubmed.ncbi.nlm.nih.gov/29075426/)]
3. Brancu M, Mann-Wrobel M, Beckham JC, Wagner HR, Elliott A, Robbins AT, et al. Subthreshold posttraumatic stress disorder: a meta-analytic review of DSM-IV prevalence and a proposed DSM-5 approach to measurement. *Psychol Trauma* 2016 Mar;8(2):222-232. [doi: [10.1037/tra0000078](https://doi.org/10.1037/tra0000078)] [Medline: [26390108](https://pubmed.ncbi.nlm.nih.gov/26390108/)]
4. Kantor V, Knepfel M, Lueger-Schuster B. Perceived barriers and facilitators of mental health service utilization in adult trauma survivors: a systematic review. *Clin Psychol Rev* 2017 Mar;52:52-68 [FREE Full text] [doi: [10.1016/j.cpr.2016.12.001](https://doi.org/10.1016/j.cpr.2016.12.001)] [Medline: [28013081](https://pubmed.ncbi.nlm.nih.gov/28013081/)]
5. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. *Mil Med* 2014 Jan;179(1):12-18. [doi: [10.7205/MILMED-D-13-00271](https://doi.org/10.7205/MILMED-D-13-00271)] [Medline: [24402979](https://pubmed.ncbi.nlm.nih.gov/24402979/)]
6. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. *J Consult Clin Psychol* 2017 Mar;85(3):267-273. [doi: [10.1037/ccp0000163](https://doi.org/10.1037/ccp0000163)] [Medline: [28221061](https://pubmed.ncbi.nlm.nih.gov/28221061/)]
7. Kuhn E, van der Meer C, Owen JE, Hoffman JE, Cash R, Carrese P, et al. PTSD Coach around the world. *Mhealth* 2018;4:15 [FREE Full text] [doi: [10.21037/mhealth.2018.05.01](https://doi.org/10.21037/mhealth.2018.05.01)] [Medline: [29963560](https://pubmed.ncbi.nlm.nih.gov/29963560/)]
8. Wickersham A, Petrides PM, Williamson V, Leightley D. Efficacy of mobile application interventions for the treatment of post-traumatic stress disorder: a systematic review. *Digit Health* 2019;5:2055207619842986 [FREE Full text] [doi: [10.1177/2055207619842986](https://doi.org/10.1177/2055207619842986)] [Medline: [31019722](https://pubmed.ncbi.nlm.nih.gov/31019722/)]
9. Krausz M, Ward J, Ramsey D. From telehealth to an interactive virtual clinic. In: *e-Mental Health*. Cham, Switzerland: Springer International Publishing; 2016.
10. Tiet QQ, Duong H, Davis L, French R, Smith CL, Leyva YE, et al. PTSD coach mobile application with brief telephone support: a pilot study. *Psychol Serv* 2019 May;16(2):227-232. [doi: [10.1037/ser0000245](https://doi.org/10.1037/ser0000245)] [Medline: [30407056](https://pubmed.ncbi.nlm.nih.gov/30407056/)]
11. Keen SM, Roberts N. Preliminary evidence for the use and efficacy of mobile health applications in managing posttraumatic stress disorder symptoms. *Health Systems* 2017 Dec 19;6(2):122-129. [doi: [10.1057/hs.2016.2](https://doi.org/10.1057/hs.2016.2)]
12. Cernvall M, Sveen J, Bergh Johannesson K, Arnberg F. A pilot study of user satisfaction and perceived helpfulness of the Swedish version of the mobile app PTSD Coach. *Eur J Psychotraumatol* 2018;9(Suppl 1):1472990 [FREE Full text] [doi: [10.1080/20008198.2018.1472990](https://doi.org/10.1080/20008198.2018.1472990)] [Medline: [29805783](https://pubmed.ncbi.nlm.nih.gov/29805783/)]
13. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: a pilot randomized controlled trial with community trauma survivors. *Psychol Trauma* 2016 May;8(3):384-392. [doi: [10.1037/tra0000092](https://doi.org/10.1037/tra0000092)] [Medline: [27046668](https://pubmed.ncbi.nlm.nih.gov/27046668/)]
14. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. *Gen Hosp Psychiatry* 2016;38:94-98. [doi: [10.1016/j.genhosppsych.2015.09.005](https://doi.org/10.1016/j.genhosppsych.2015.09.005)] [Medline: [26589765](https://pubmed.ncbi.nlm.nih.gov/26589765/)]
15. Pacella-LaBarbara ML, Suffoletto BP, Kuhn E, Germain A, Jaramillo S, Repine M, et al. A pilot randomized controlled trial of the PTSD coach app following motor vehicle crash-related injury. *Acad Emerg Med* 2020 Nov 19;27(11):1126-1139 [FREE Full text] [doi: [10.1111/acem.14000](https://doi.org/10.1111/acem.14000)] [Medline: [32339359](https://pubmed.ncbi.nlm.nih.gov/32339359/)]
16. Shakespeare-Finch J, Alichniewicz KK, Strodl E, Brown K, Quinn C, Hides L, et al. Experiences of serving and ex-serving members with the PTSD coach Australia app: mixed methods study. *J Med Internet Res* 2020 Oct 08;22(10):e18447 [FREE Full text] [doi: [10.2196/18447](https://doi.org/10.2196/18447)] [Medline: [33030438](https://pubmed.ncbi.nlm.nih.gov/33030438/)]
17. Rozental A, Andersson G, Boettcher J, Ebert DD, Cuijpers P, Knaevelsrud C, et al. Consensus statement on defining and measuring negative effects of internet interventions. *Internet Interventions* 2014 Mar;1(1):12-19. [doi: [10.1016/j.invent.2014.02.001](https://doi.org/10.1016/j.invent.2014.02.001)]
18. Linden M. How to define, find and classify side effects in psychotherapy: from unwanted events to adverse treatment reactions. *Clin Psychol Psychother* 2013 Jan 18;20(4):286-296. [doi: [10.1002/cpp.1765](https://doi.org/10.1002/cpp.1765)] [Medline: [22253218](https://pubmed.ncbi.nlm.nih.gov/22253218/)]
19. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. *JMIR Ment Health* 2015;2(1):e7 [FREE Full text] [doi: [10.2196/mental.3935](https://doi.org/10.2196/mental.3935)] [Medline: [26543913](https://pubmed.ncbi.nlm.nih.gov/26543913/)]
20. Pacella ML, Hruska B, Delahanty DL. The physical health consequences of PTSD and PTSD symptoms: a meta-analytic review. *J Anxiety Disord* 2013 Jan;27(1):33-46. [doi: [10.1016/j.janxdis.2012.08.004](https://doi.org/10.1016/j.janxdis.2012.08.004)] [Medline: [23247200](https://pubmed.ncbi.nlm.nih.gov/23247200/)]
21. Gupta MA. Review of somatic symptoms in post-traumatic stress disorder. *Int Rev Psychiatry* 2013 Feb;25(1):86-99. [doi: [10.3109/09540261.2012.736367](https://doi.org/10.3109/09540261.2012.736367)] [Medline: [23383670](https://pubmed.ncbi.nlm.nih.gov/23383670/)]
22. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 2007 May;39(2):175-191. [doi: [10.3758/BF03193146](https://doi.org/10.3758/BF03193146)]

23. 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](https://doi.org/10.1016/j.jbi.2008.08.010)] [Medline: [18929686](https://pubmed.ncbi.nlm.nih.gov/18929686/)]
24. Hensler I, Sveen J, Cernvall M, Arnberg FK. Ecological momentary assessment of self-rated health, daily strategies and self-management app use among trauma-exposed adults. *Eur J Psychotraumatol* 2021 May 27;12(1):1920204 [FREE Full text] [doi: [10.1080/20008198.2021.1920204](https://doi.org/10.1080/20008198.2021.1920204)] [Medline: [34104354](https://pubmed.ncbi.nlm.nih.gov/34104354/)]
25. PTSD: national center for PTSD. U.S. Department of Veterans Affairs. URL: https://www.ptsd.va.gov/appvid/mobile/ptsdcoach_app.asp [accessed 2022-02-15]
26. 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](https://doi.org/10.1037/pas0000254)] [Medline: [26653052](https://pubmed.ncbi.nlm.nih.gov/26653052/)]
27. Bondjers K. Post-traumatic Stress Disorder – Assessment of Current Diagnostic Definitions. Uppsala, Sweden: Acta Universitatis Upsaliensis; 2020. URL: <http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-403118>
28. Sveen J, Bondjers K, Willebrand M. Psychometric properties of the PTSD Checklist for DSM-5: a pilot study. *Eur J Psychotraumatol* 2016;7:30165 [FREE Full text] [doi: [10.3402/ejpt.v7.30165](https://doi.org/10.3402/ejpt.v7.30165)] [Medline: [27098450](https://pubmed.ncbi.nlm.nih.gov/27098450/)]
29. PTSD checklist for DSM-5 (PCL-5). U.S. Department of Veterans Affairs. URL: <https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp> [accessed 2022-02-15]
30. Monson CM, Gradus JL, Young-Xu Y, Schnurr PP, Price JL, Schumm JA. Change in posttraumatic stress disorder symptoms: do clinicians and patients agree? *Psychol Assess* 2008 Jun;20(2):131-138. [doi: [10.1037/1040-3590.20.2.131](https://doi.org/10.1037/1040-3590.20.2.131)] [Medline: [18557690](https://pubmed.ncbi.nlm.nih.gov/18557690/)]
31. Hansson M, Chotai J, Nordstöm A, Bodlund O. Comparison of two self-rating scales to detect depression: HADS and PHQ-9. *Br J Gen Pract* 2009 Sep 01;59(566):e283-e288. [doi: [10.3399/bjgp09x454070](https://doi.org/10.3399/bjgp09x454070)]
32. 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](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
33. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. *J Gen Intern Med* 2007 Nov;22(11):1596-1602 [FREE Full text] [doi: [10.1007/s11606-007-0333-y](https://doi.org/10.1007/s11606-007-0333-y)] [Medline: [17874169](https://pubmed.ncbi.nlm.nih.gov/17874169/)]
34. Nordin S, Palmquist E, Nordin M. Psychometric evaluation and normative data for a Swedish version of the Patient Health Questionnaire 15-Item Somatic Symptom Severity Scale. *Scand J Psychol* 2013 Apr;54(2):112-117. [doi: [10.1111/sjop.12029](https://doi.org/10.1111/sjop.12029)] [Medline: [23294182](https://pubmed.ncbi.nlm.nih.gov/23294182/)]
35. Kroenke K, Spitzer RL, Williams JB. The PHQ-15: validity of a new measure for evaluating the severity of somatic symptoms. *Psychosom Med* 2002;64(2):258-266. [doi: [10.1097/00006842-200203000-00008](https://doi.org/10.1097/00006842-200203000-00008)] [Medline: [11914441](https://pubmed.ncbi.nlm.nih.gov/11914441/)]
36. Rozental A, Kottorp A, Forsström D, Månsson K, Boettcher J, Andersson G, et al. The Negative Effects Questionnaire: psychometric properties of an instrument for assessing negative effects in psychological treatments. *Behav Cognit Psychother* 2019 Mar 15;47(5):559-572. [doi: [10.1017/s1352465819000018](https://doi.org/10.1017/s1352465819000018)]
37. Life Events Checklist for DSM-5 (LEC-5). U.S. Department of Veterans Affairs. URL: https://www.ptsd.va.gov/professional/assessment/te-measures/life_events_checklist.asp [accessed 2022-02-15]
38. Lecrubier Y, Sheehan D, Weiller E, Amorim P, Bonora I, Sheehan K, et al. The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: reliability and validity according to the CIDI. *Eur Psychiatr* 2020 Apr 16;12(5):224-231. [doi: [10.1016/S0924-9338\(97\)83296-8](https://doi.org/10.1016/S0924-9338(97)83296-8)]
39. Applied missing data analysis with SPSS and (R)Studio. BookDown. 2019. URL: <https://bookdown.org/mwheymans/bookmi/> [accessed 2022-02-15]
40. Feingold A. Confidence interval estimation for standardized effect sizes in multilevel and latent growth modeling. *J Consult Clin Psychol* 2015 Feb;83(1):157-168 [FREE Full text] [doi: [10.1037/a0037721](https://doi.org/10.1037/a0037721)] [Medline: [25181028](https://pubmed.ncbi.nlm.nih.gov/25181028/)]
41. Torous J, Firth J. The digital placebo effect: mobile mental health meets clinical psychiatry. *Lancet Psychiatry* 2016 Feb;3(2):100-102. [doi: [10.1016/S2215-0366\(15\)00565-9](https://doi.org/10.1016/S2215-0366(15)00565-9)]
42. Households in Sweden [Swedish]. Statistics Sweden. URL: <http://www.scb.se/hitta-statistik/sverige-i-siffror/manniskorna-i-sverige/hushall-i-sverige/> [accessed 2022-02-15]
43. The level of education in Sweden [Swedish]. Statistics Sweden. URL: <http://www.scb.se/hitta-statistik/sverige-i-siffror/utbildning-jobb-och-pengar/utbildningsnivan-i-sverige/> [accessed 2022-02-15]
44. Employment in Sweden [Swedish]. Statistics Sweden. URL: <http://www.scb.se/hitta-statistik/sverige-i-siffror/samhallets-ekonomi/sysselsattning-i-sverige/> [accessed 2022-02-15]
45. PTSD Coach Sweden. OSF HOME. URL: <https://doi.org/10.17605/OSF.IO/F6P3R> [accessed 2022-02-15]

Abbreviations

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

PCL-5: Posttraumatic Stress Disorder Checklist for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition

PHQ: Patient Health Questionnaire
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
REDCap: Research Electronic Data Capture

Edited by A Mavragani; submitted 22.06.21; peer-reviewed by S Woodward, A De la Rosa-Gómez, D Hasson; comments to author 25.08.21; revised version received 19.12.21; accepted 05.01.22; published 30.03.22.

Please cite as:

Hensler I, Sveen J, Cernvall M, Arnberg FK

Efficacy, Benefits, and Harms of a Self-management App in a Swedish Trauma-Exposed Community Sample (PTSD Coach): Randomized Controlled Trial

J Med Internet Res 2022;24(3):e31419

URL: <https://www.jmir.org/2022/3/e31419>

doi: [10.2196/31419](https://doi.org/10.2196/31419)

PMID: [35353052](https://pubmed.ncbi.nlm.nih.gov/35353052/)

©Ida Hensler, Josefin Sveen, Martin Cernvall, Filip K Arnberg. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 30.03.2022. 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.

Original Paper

Effect of a Lifestyle-Focused Web-Based Application on Risk Factor Management in Patients Who Have Had a Myocardial Infarction: Randomized Controlled Trial

Halldóra Ögmundsdóttir Michelsen^{1,2}, PhD; Ingela Sjölin³, MA; Maria Bäck^{4,5}, PhD; Manuel Gonzalez Garcia^{6,7}, PhD; Anneli Olsson^{8,9}, MSc; Camilla Sandberg^{6,10}, PhD; Alexandru Schiopu^{1,11}, PhD; Margrét Leósdóttir^{1,3}, PhD

¹Department of Clinical Sciences Malmö, Lund University, Malmö, Sweden

²Department of Internal Medicine, Helsingborg Hospital, Helsingborg, Sweden

³Department of Cardiology, Skane University Hospital, Malmö, Sweden

⁴Department of Health, Medicine and Caring Sciences, Unit of Physiotherapy, Linköping University, Linköping, Sweden

⁵Department of Occupational Therapy and Physiotherapy, Sahlgrenska University Hospital, Gothenburg, Sweden

⁶Department of Epidemiology and Global Health, Umeå University, Umeå, Sweden

⁷Faculty of Medicine, The University of Queensland, Herston, Queensland, Australia

⁸Department of Cardiology, Skane University Hospital, Lund, Sweden

⁹Department of Clinical Sciences, Lund University, Lund, Sweden

¹⁰Department of Community Medicine and Rehabilitation, Physiotherapy, Umeå University, Umeå, Sweden

¹¹Department of Internal Medicine, Skane University Hospital, Lund, Sweden

Corresponding Author:

Halldóra Ögmundsdóttir Michelsen, PhD

Department of Clinical Sciences Malmö

Lund University

Jan Waldenströms Gata 35

Malmö, 205 02

Sweden

Phone: 46 4061000

Email: halldora.michelsen@med.lu.se

Abstract

Background: Cardiac rehabilitation is central in reducing mortality and morbidity after myocardial infarction. However, the fulfillment of guideline-recommended cardiac rehabilitation targets is unsatisfactory. eHealth offers new possibilities to improve clinical care.

Objective: This study aims to assess the effect of a web-based application designed to support adherence to lifestyle advice and self-control of risk factors (intervention) in addition to center-based cardiac rehabilitation, compared with cardiac rehabilitation only (usual care).

Methods: All 150 patients participated in cardiac rehabilitation. Patients randomized to the intervention group (n=101) received access to the application for 25 weeks where information about lifestyle (eg, diet and physical activity), risk factors (eg, weight and blood pressure [BP]), and symptoms could be registered. The software provided feedback and lifestyle advice. The primary outcome was a change in submaximal exercise capacity (Watts [W]) between follow-up visits. Secondary outcomes included changes in modifiable risk factors between baseline and follow-up visits and uptake and adherence to the application. Regression analysis was used, adjusting for relevant baseline variables.

Results: There was a nonsignificant trend toward a larger change in exercise capacity in the intervention group (n=66) compared with the usual care group (n=40; +14.4, SD 19.0 W, vs +10.3, SD 16.1 W; $P=.22$). Patients in the intervention group achieved significantly larger BP reduction compared with usual care patients at 2 weeks (systolic -27.7 vs -16.4 mm Hg; $P=.006$) and at 6 to 10 weeks (systolic -25.3 vs -16.4 mm Hg; $P=.02$, and diastolic -13.4 vs -9.1 mm Hg; $P=.05$). A healthy diet index score improved significantly more between baseline and the 2-week follow-up in the intervention group (+2.3 vs +1.4 points; $P=.05$), mostly owing to an increase in the consumption of fish and fruit. At 6 to 10 weeks, 64% (14/22) versus 46% (5/11) of smokers in the intervention versus usual care groups had quit smoking, and at 12 to 14 months, the respective percentages were 55%

(12/22) versus 36% (4/11). However, the number of smokers in the study was low (33/149, 21.9%), and the differences were nonsignificant. Attendance in cardiac rehabilitation was high, with 96% (96/100) of patients in the intervention group and 98% (48/49) of patients receiving usual care only attending 12- to 14-month follow-up. Uptake (logging data in the application at least once) was 86.1% (87/101). Adherence (logging data at least twice weekly) was 91% (79/87) in week 1 and 56% (49/87) in week 25.

Conclusions: Complementing cardiac rehabilitation with a web-based application improved BP and dietary habits during the first months after myocardial infarction. A nonsignificant tendency toward better exercise capacity and higher smoking cessation rates was observed. Although the study group was small, these positive trends support further development of eHealth in cardiac rehabilitation.

Trial Registration: ClinicalTrials.gov NCT03260582; <https://clinicaltrials.gov/ct2/show/NCT03260582>

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-018-3118-1

(*J Med Internet Res* 2022;24(3):e25224) doi:[10.2196/25224](https://doi.org/10.2196/25224)

KEYWORDS

eHealth; cardiac rehabilitation; cardiovascular; mobile device app; risk factors; web-based application; mobile phone

Introduction

Background

Mortality rates from coronary heart disease (CHD) have decreased in the last decades [1-3]. However, the falling mortality rates have led to an increased number of survivors who need support in the secondary prevention of recurrent coronary events [3,4]. Comprehensive cardiac rehabilitation programs, which are multidisciplinary medical and health behavioral interventions, effectively reduce CHD morbidity and mortality [3-7]. International guidelines strongly recommend participation in cardiac rehabilitation and have set therapeutic goals on risk factor and lifestyle management [7]. Cardiac rehabilitation goals include smoking cessation, medical management of low-density lipoprotein (LDL)-cholesterol and blood pressure (BP), maintaining a healthy diet, and participation in an exercise training program, also referred to as exercise-based cardiac rehabilitation [7]. Despite clear set goals, both attendance to cardiac rehabilitation and therapeutic goal attainment are suboptimal [8,9].

Cardiac rehabilitation programs in Sweden hold a high standard compared with European counterparts when it comes to providing guideline-recommended services [10]. Moreover, in 2019, the Swedish Web-System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies (SWEDEHEART) registry reported that after participating in a 1-year long cardiac rehabilitation program, patients who had an myocardial infarction (MI) only reached target levels for LDL-cholesterol in about 60% of cases, only 55% of smokers were abstinent and 19% of patients had participated in exercise-based cardiac rehabilitation programs [9]. A lack of reaching therapeutic goals while offering guideline-recommended services indicates that alternative methods may be needed to improve patient outcomes.

Previous studies on eHealth cardiac rehabilitation have resulted in noninferiority or concluded that patients benefited with regard to lifestyle changes and risk factor management [11-20]. However, these studies have varied in sample size and follow-up time, and more studies are needed. eHealth is of interest to the

field of cardiac rehabilitation because it has the potential to overcome some known barriers to participation, such as geographical distance, communication barriers, and rigid follow-up structures, and to individualize cardiac rehabilitation programs [4,21].

Objectives

The aim of this study is to evaluate the efficacy of a web-based application as an addition to a comprehensive center-based cardiac rehabilitation program, in comparison with usual care center-based cardiac rehabilitation only. The web-based application was designed to support adherence to lifestyle advice and self-control of risk factors in patients who had an MI. In the planning phase of our study, we predicted that the web-based application would primarily affect patients' lifestyle and with that increased physical activity levels and exercise capacity [22]. The benefits of increased physical activity levels and participation in exercise-based cardiac rehabilitation are largely mediated through an increase in physical fitness. Submaximal exercise capacity is an objective measurement of physical fitness and was, therefore, chosen as the primary outcome [23].

Methods

Trial Design and Framework

The protocol, which followed the principles outlined in the Declaration of Helsinki, was approved by the Regional Ethical Review Board at Lund University (approval 2016/5). The protocol has been previously described in detail [22] and is summarized here.

We conducted an unblinded parallel multicenter randomized controlled trial (RCT) at 3 cardiac rehabilitation centers based at university hospitals in Sweden. At the time of the study, approximately 1200 patients aged <75 years were treated for acute MI at the 3 study centers each year.

Participants, Recruitment, and Randomization

The inclusion criteria were age 18-74 years, having had an MI within the last 2 weeks, owning a smartphone or having access to the internet via a computer or tablet, and being able to handle the software. The exclusion criteria were having an expected

survival of <1 year, dementia, severe psychiatric illness or drug abuse, severe physical disability limiting the patients' ability to participate in a center-based exercise-based cardiac rehabilitation, inability to speak or understand the Swedish language, and a 3-vessel or left main disease requiring coronary artery bypass grafting. The age criterion was set to match that applied by the SWEDEHEART registry.

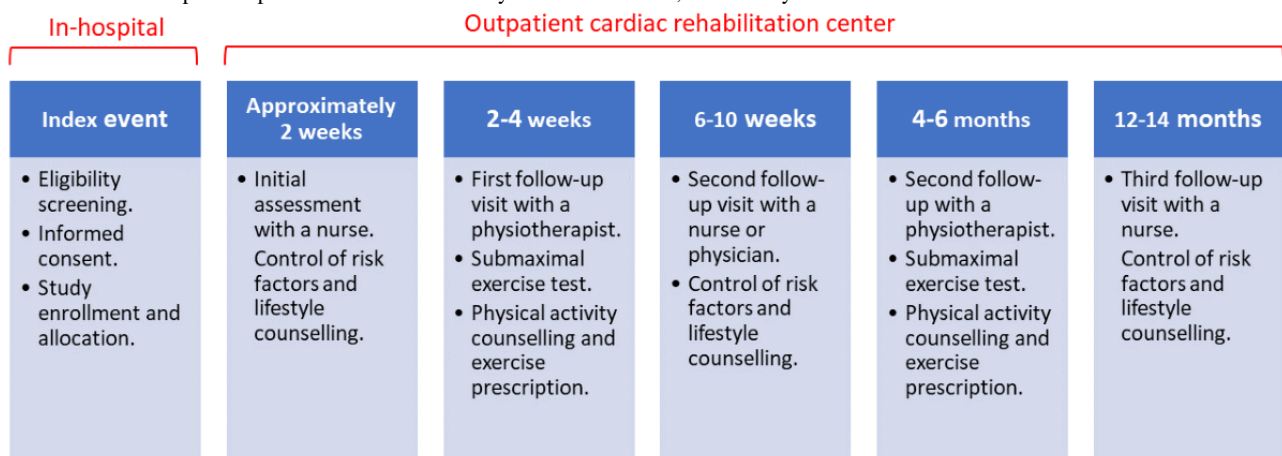
Eligibility screening and study inclusion were performed within 2 weeks of an index MI while participants were admitted to a coronary care unit at each participating hospital. Local study coordinators (physicians, nurses, or physiotherapists) provided eligible patients with information about the study, offered participation, and obtained written informed consent. Once included in the study, the participants were randomized to 1 of the 2 study arms using opaque sealed envelopes. The envelopes,

which included information on which study group the participant would be randomized to, were prepared by a member of the research team and shuffled by another member. The envelopes were thereafter sequentially numbered with unique numbers for each recruiting site. Upon recruitment, baseline questionnaires were administered.

Usual Care

Participants in both arms of the study were offered participation in a comprehensive cardiac rehabilitation program at each center. A description of the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) figure on the content of each follow-up visit is available in the protocol [22]. In short, the cardiac rehabilitation programs consisted of five outpatient follow-up visits: 3 visits with a nurse or physician and 2 visits with a physiotherapist (Figure 1).

Figure 1. The follow-up of the patients who have had a myocardial infarction, in the study.



The Intervention

Participants randomized to the intervention group of the study received access to the LifePod web-based mobile device app (Cross Technology Solutions AB) for the first 25 weeks of the cardiac rehabilitation program. The LifePod software contained two separate interfaces: one for the patient and one for the treating health care professionals (Figure 2). In the patient interface, the patient could log information about diet, physical activity and exercise, weight, heart rate, BP, and smoking, as

well as symptoms and intake of medication. The patient could compare his or her own data to guideline-recommended targets and received automated positive feedback on healthy lifestyle choices as well as general recommendations on exercise, daily physical activity, and healthy diet. In the medical interface assessed by the treating cardiac rehabilitation staff, the system ranked patients, giving high priority to, for example, patients reporting chest pain or out-of-range BP measurements. The medical interface was reviewed twice a week by a nurse.

Figure 2. Screenshots of the LifePod patient interface (top) and medical interface (bottom). The software could be accessed through a smartphone, tablet, or computer.



Data Collection

Patient outcomes were derived from the SWEDHEART registry. The SWEDHEART registry is a nationwide quality registry that records baseline characteristics, treatments, follow-up, and outcome data of patients who have had an MI [9]. Data collection started at the time of index MI and continued at all follow-up visits, as shown in Figure 1. The time points for data collection were prespecified in accordance with standardized SWEDHEART registry follow-up visits. Information was collected using standardized forms that can be downloaded from the SWEDHEART registry webpage [24].

At the physiotherapist visits at 2 to 4 weeks and 4 to 6 months, a submaximal exercise test on a bicycle ergometer was performed [25,26]. During the test, heart rate, systolic BP (SBP), perceived exertion, dyspnea, and chest pain were registered according to the Borg rating of perceived exertion and category ratio-10 scales [27]. The test was discontinued at Borg rating of perceived exertion 17, dyspnea 7 on the Borg category ratio-10 scale, or other routine discontinuation criteria for exercise stress testing, including, for example, chest pain and fall in SBP. At the first visit, the patients received individualized physical activity and exercise recommendations and were invited

to participate in exercise-based cardiac rehabilitation as a part of their comprehensive cardiac rehabilitation program.

The follow-up visits with a nurse or physician at approximately 2 weeks, 6 to 10 weeks, and 12 to 14 months focused on risk factors and lifestyle. Fasting plasma glucose (mmol/L), hemoglobin A_{1c} (mmol/mol), and plasma lipids (total cholesterol, LDL-cholesterol, high-density lipoprotein-cholesterol, and triglycerides; mmol/L) samples were drawn and analyzed using accredited methods at each hospital. BP was measured using a manual sphygmomanometer after a 5-minute rest with the patient in a sitting position. Weight (kg) was measured in light indoor clothing, and BMI (kg/m²) was calculated. Smoking status was self-reported. At baseline, 6 to 10 weeks, and 12 to 14 months, smoking abstinence was defined as being smoke-free for ≥1 month. At the 2-week visit, abstinence was defined as being abstinent at the time of the visit. Diet was evaluated using a 4-item questionnaire adapted from the national guidelines for the management of unhealthy lifestyle in the general population [28]. The questions aim to quantify the amount of vegetables, fruit, fish, and sweets consumed. Each question had 4 possible answers, giving 0 to 3 points. The scores for each question were subsequently added, forming the healthy diet index (0-12 points). Levels of physical

activity were self-reported using 2 sets of questionnaires. Haskell questions on physical activity and exercise evaluated both the number of days during the last week [1-7] with at least 30 minutes of physical activity and the number of days during the last week [1-7] with at least 20 minutes of exercise training [29]. The Frändin-Grimby physical activity questionnaire aims to evaluate the level of physical activity a person achieved in the last week on a scale of 1 to 6, with 1 being hardly any physical activity and 6 being regular strenuous physical activity [30].

The data that support this study's findings are available from the Uppsala Clinical Research Centre, Sweden. The primary responsibility for data monitoring, including data sharing, integrity, and security, was assigned to the principal investigator and local study coordinators. The trial was conducted according to Good Clinical Practice, and data were handled according to the Swedish Data Protection Authority and General Data Protection Regulations.

Statistical Methods and Outcomes

With a power of 90%, a 2-sided significance level of .05, and a mean difference of at least 10 (SD 20) W among the groups, the estimated sample size needed was 50 participants in each group [22]. However, the expected loss of adherence to the web-based application led to the formation of an unequal allocation ratio of 1:2 in the usual care group versus intervention group.

The primary outcome was the change (δ value) in submaximal exercise capacity, measured in Watts, between 2 submaximal exercise tests performed at 2 to 4 weeks and at 4 to 6 months. The secondary outcomes were changes (δ values) in dietary habits and physical activity, as well as SBP and diastolic BP (DBP), total cholesterol, LDL-cholesterol, high-density lipoprotein-cholesterol, triglycerides, fasting plasma glucose, hemoglobin A_{1c}, and BMI. The secondary outcomes also included smoking status and uptake and adherence to the web-based application. Uptake was defined as the proportion of patients who logged on to the patient interface at least once, and adherence was defined as the proportion of patients registering data at least twice per week on a weekly basis throughout the 25-week intervention period. An additional

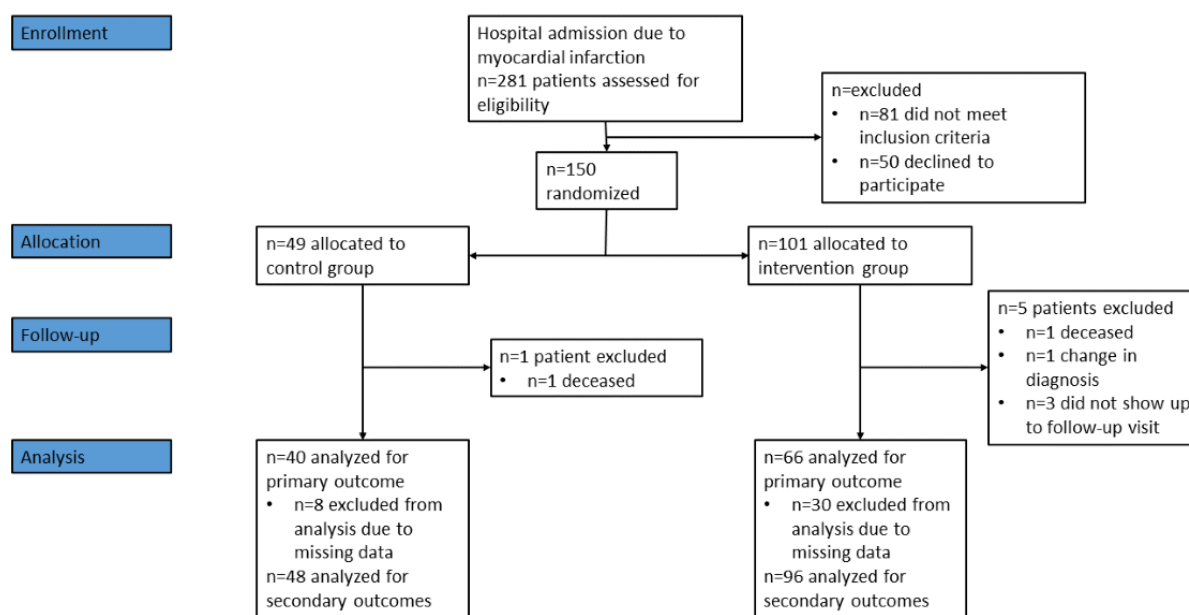
analysis on the difference in mortality and number of hospital readmissions during the trial period was performed. Baseline characteristics are presented as means (SDs) for normally distributed continuous variables, medians (IQR) for nonnormally distributed continuous variables, and as numbers and percentages for categorical variables. Variable distribution was assessed by visual inspection of histograms and Q-Q plots and by calculating skewness and kurtosis; z values between -1.96 and 1.96 were used to define the normally distributed interval. To compare outcome measures of continuous variables among the groups, a 2-tailed Student t test or a Mann-Whitney U test was performed. A univariate analysis of variance, adjusting for age, gender, weight, previous CHD, and smoking status at index MI was also performed. In the analysis of variance, the dependent variables were the measured outcomes, and the independent variables were the intervention and chosen covariates. For categorical variables, a chi-square test and a logistic regression analysis adjusting for the previously mentioned covariates were used. For within-group comparisons, a paired t test was used for normally distributed variables and a Wilcoxon signed-rank test for skewed data. For all analyses, an α level of .05, and 2-tailed testing was used. Individuals with missing data on covariates or outcome variables were excluded from the analysis (listwise exclusion). All data were analyzed by using the SPSS (version 25.0; IBM Corp).

Results

Participant Recruitment and Flow

Of the 281 patients assessed for eligibility, 150 (53.4%) consented and indicated intent to attend cardiac rehabilitation and, if allocated to the intervention group, use the web-based application (Figure 3). In total, 32.7% (49/150) of the patients were allocated to the control group and 67.3% (101/150) to the intervention group. At the end of the study period, 1 participant in each group of the study had died. For 1 patient, the index event diagnosis changed during follow-up from MI to cardiac amyloidosis, and the patient was excluded from the analysis. Of the 150 patients, 3 (2%) did not attend their follow-up visits at 12 to 14 months after MI and were classified as lost to follow-up.

Figure 3. A flowchart displaying the recruitment process and flow of participants in the usual care and intervention groups.



Patient inclusion started in April 2016 and was finalized in April 2018. The follow-up period was completed in June 2019.

Baseline Characteristics

Baseline characteristics are presented in Table 1. The intervention group had a slightly higher prevalence of cardiac

comorbidities at baseline including a higher mean SBP and were more often taking cardioprotective medication compared with the usual care group. In addition, a proportionally higher number of patients in the intervention group had an ST elevation MI.

Table 1. Baseline characteristics.^a

Parameters	Intervention (n=100)	Usual care (n=49)
Age (years), mean (SD)	60.0 (8.9)	61.1 (8.6)
Male gender, n (%)	84 (84)	36 (73)
Active smoker, n (%)	22 (22)	11 (22)
Physiological measures		
SBP ^b (mm Hg), mean (SD)	150.0 (27.6)	142.9 (25.5)
DBP ^c (mm Hg), mean (SD)	88.5 (14.6)	86.8 (14.8)
Waist circumference (cm), mean (SD)	104.9 (12.9)	104.5 (13.9)
Weight (kg), mean (SD)	86.3 (15.1)	85.3 (16.2)
BMI (kg/m ²), median (IQR)	27 (25-30)	27 (25-29)
Laboratory measures		
Total cholesterol (mmol/L), mean (SD)	4.7 (1.1)	4.9 (1.1)
LDL ^d -cholesterol (mmol/L), mean (SD)	2.8 (0.9)	3 (1.0)
Triglycerides (mmol/L), median (IQR)	1.4 (0.9-2.1)	1.4 (1.0-1.9)
HDL ^e -cholesterol (mmol/L), median (IQR)	1.2 (0.9-1.4)	1.2 (0.9-1.4)
Fasting plasma glucose (mmol/L), median (IQR)	7.5 (6.4-9.2)	7.1 (6.2-8.9)
HbA _{1c} ^f (mmol/mol), median (IQR)	38 (34.5-41.0)	39 (36.0-42.0)
Previous disease, n (%)		
Ischemic heart disease (previous MI ^g , PCI ^h , or CABG ⁱ)	18 (18)	6 (12)
Heart failure	5 (5)	0 (0)
Diabetes mellitus	9 (9)	6 (12)
Hypertension	42 (42)	17 (34)
Medication on hospital admission, n (%)		
ACEi ^j or ARB ^k	36 (36)	12 (24)
Statins	24 (24)	5 (10)
Acetylsalicylic acid	18 (18)	5 (10)
β-blockers	19 (19)	7 (14)
Medication at hospital discharge, n (%)		
ACEi or ARB	95 (95)	46 (93)
Statins	100 (100)	47 (95)
DAPT ^l	100 (100)	49 (100)
β-blockers	89 (89)	43 (87)
Type of MI, n (%)		
STEMI ^m	59 (59)	24 (48)
NSTEMI ⁿ	41 (41)	24 (48)

^aData are presented as n (%), mean (SD), or median (IQR).

^bSBP: systolic blood pressure.

^cDBP: diastolic blood pressure.

^dLDL: low-density lipoprotein.

^eHDL: high-density lipoprotein.

^fHbA_{1c}: glycated hemoglobin A_{1c}.

^gMI: myocardial infarction.

^hPCI: percutaneous coronary intervention.

ⁱCABG: coronary artery bypass grafting.

^jACEi: angiotensin-converting enzyme inhibitor.

^kARB: angiotensin II receptor blocker.

^lDAPT: dual antiplatelet therapy.

^mSTEMI: ST elevation myocardial infarction.

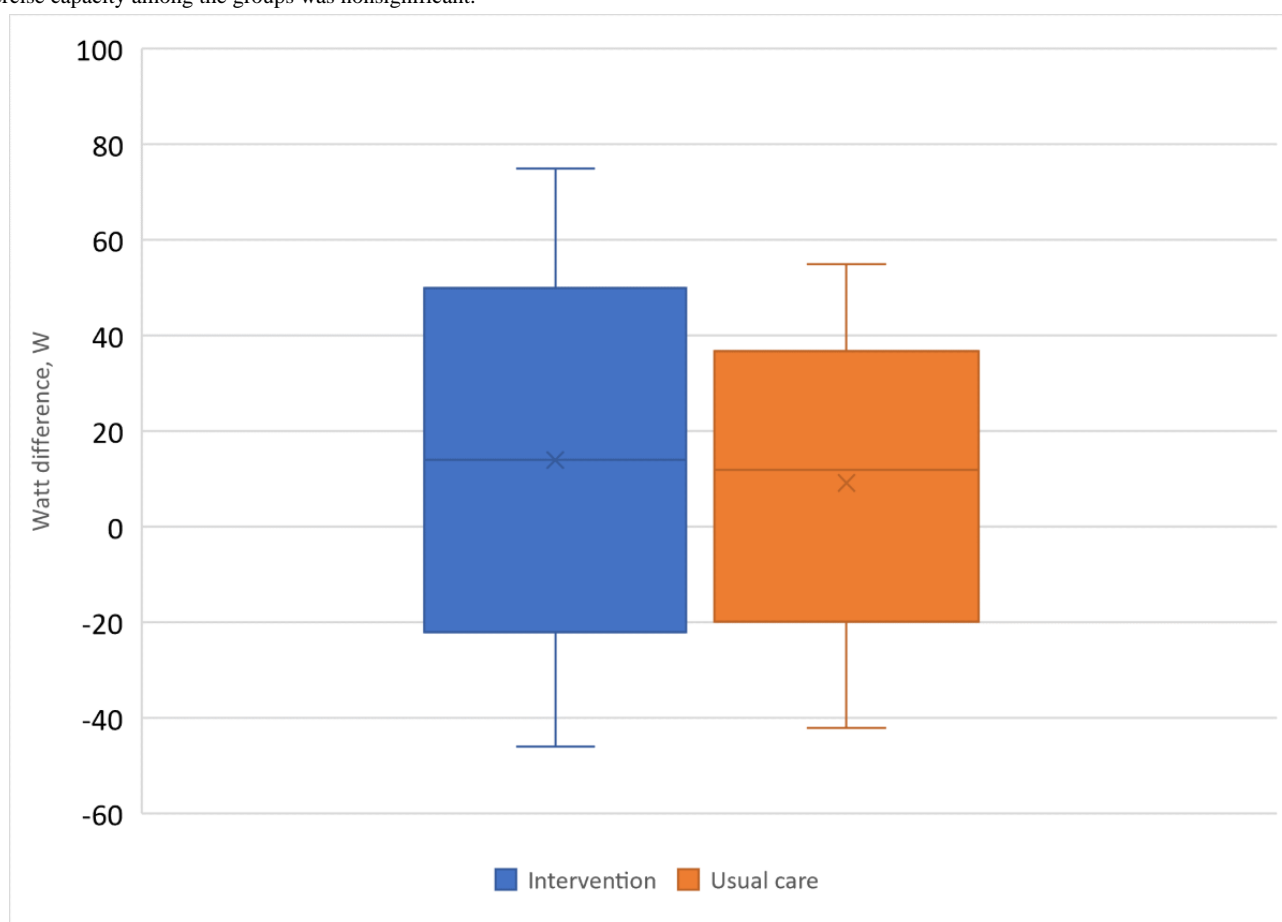
ⁿNSTEMI: non-ST elevation myocardial infarction.

Primary Outcome

Because of missing data at the first or second follow-up with a physiotherapist, only 82% (40/49) of participants in the usual care group and 66% (66/100) in the intervention group qualified for analyses of the primary outcome.

Both groups increased their submaximal exercise capacity significantly between the first and second follow-up measurements (Figure 4). The intervention group increased their exercise capacity from 96.3 (SD 29.4) W to 110.8 W (SD 33.7 W; $P<.001$) and the corresponding values for the usual care group were 96.1 (SD 33.7) W to 106.5 W (SD 37.3 W; $P<.001$).

Figure 4. A boxplot of the change in submaximal exercise capacity in the intervention group and the usual care group. The difference in change in exercise capacity among the groups was nonsignificant.



There was no significant difference in the change in exercise capacity between the intervention group and the usual care group (+14.4, SD 19.0 W vs +10.3, SD 16.1 W; crude 95% CI -11.2 to 3.0, $P=.26$; adjusted 95% CI -12.0 to 2.8, $P=.22$) at follow-up. The variation in the observed difference fell within the expected limits of +20 W to -20 W.

Secondary Outcomes

The secondary outcomes are shown in Tables 2-4. Crude analyses showed that at the 2-week follow-up, there was a

significantly larger decrease in SBP in the intervention group than that in the usual care group. SBP remained lower in the intervention group than in the usual care group for the remainder of the follow-ups; however, the difference among the groups was nonsignificant. In the adjusted analysis, the decrease in SBP was significantly larger in the intervention group between baseline and 2 weeks and for both SBP and DBP between baseline and 6 to 10 weeks (Figure 5).

Table 2. Secondary outcome measures at the 2-week follow-up.^a

Characteristics	Intervention		Usual care		<i>P</i> value for difference crude	<i>P</i> value for difference adjusted ^b
	Values	Change from baseline	Values	Change from baseline		
Risk factors						
SBP ^c (mm Hg), mean (SD)	123.4 (17.4)	-27.7 (27.6)	127.3 (13.9)	-16.4 (24.1)	.02	.006
DBP ^d (mm Hg), mean (SD)	76.0 (10.4)	-13.1 (13.0)	76.1 (10.2)	-11.1 (15.1)	.43	.40
BMI (kg/m ²), median (IQR)	27.1 (24.6 to 31.7)	-0.3 (-0.9 to 0.5)	27.8 (24.8 to 30.8)	-0.1 (-0.7 to 0.6)	.34	.77 ^e
Self-reported parameters, mean (SD)						
Vegetable consumption	2.2 (0.9)	0.4 (0.9)	2.2 (1.0)	0.4 (0.8)	.81	.94
Fruit consumption	2.3 (0.7)	0.7(1.0)	2.2 (0.9)	0.2 (1.0)	.03	.03
Fish consumption	1.9 (1.0)	0.8 (1.0)	1.7 (1.0)	0.4 (0.8)	.03	.02
Consumption of sweets	2.4 (0.8)	0.5 (0.9)	2.4 (0.7)	0.5 (0.9)	.99	.99
Healthy diet index	8.7 (2.1)	2.3 (2.1)	8.3 (2.1)	1.4 (2.3)	.05	.03

^aNumbers are presented as mean (SD), median (q1, q3), and *P* values. Crude and adjusted *P* values are shown for the differences in mean or median change (δ) between baseline and 2-week follow-up, comparing the intervention and usual care groups.

^bAdjusted for gender, age, weight, previous heart disease, and smoking status at the time of the index event.

^cSBP: systolic blood pressure.

^dDBP: diastolic blood pressure.

^eNot adjusted for weight.

The healthy diet index scores improved to a significantly larger extent between baseline and the 2-week follow-up in the intervention group than in the usual care group, mostly because of an increase in the consumption of fish and fruit (Table 2). However, this improvement was not sustained for the rest of the follow-up period (Tables 3 and 4).

There were no significant differences in self-reported physical activity among the groups. Results from the Haskell questions on physical activity and exercise [29] showed a mean change in number of days of performing at least 30 minutes of physical activity between baseline and the first follow-up with a physiotherapist at 2 to 4 weeks, which was 3.1 days (SD 2.2 days) for the intervention group and 2.5 days (SD 2.3 days) for the usual care group (adjusted *P*=.13). The corresponding numbers between baseline and the second follow-up with a physiotherapist at 4 to 6 months were 2.6 days (SD 2.6 days) and 1.8 days (SD 2.2 days) for the intervention and control groups, respectively (adjusted *P*=.08). The mean change in the number of days performing at least 20 minutes of exercise training between baseline and the first follow-up with a physiotherapist was 0.2 days (SD 2.0 days) for the intervention group and 0.1 days (SD 1.7 days) for the control group (adjusted

P=.76), and the corresponding numbers between baseline and the second follow-up were 1.2 days (SD 2.2 days) and 1.3 days (SD 2.4 days) for the intervention and control groups, respectively (adjusted *P*=.79).

According to the Frändin-Grimby physical activity questionnaire [30], the mean change in the level of physical activity a person achieved in the last week between baseline and the first follow-up with a physiotherapist at 2 to 4 weeks was 0.1 (SD 1.0) point in the intervention group and 0.3 (SD 1.0) point for the usual care group (adjusted *P*=.21). The mean change between baseline and the second follow-up visit with a physiotherapist at 4- to 6-month visit was 0.6 (SD 1.2) in the intervention group and 0.8 (SD 1.1) in the usual care group (adjusted *P*=.46).

A total of 33 smokers were included in the trial, 22 in the intervention group and 11 in the usual care group. At the 2-week follow-up, 64% (14/22) of smokers in the intervention group and 55% (6/11) of smokers in the usual care group reported being abstinent from smoking (adjusted *P*=.76). At 6 to 10 weeks, the numbers were 64% (14/22) vs 46% (5/11; adjusted *P*=.24), and at 12 to 14 months, they were 55% (12/22) vs 36% (4/11; adjusted *P*=.74), for the intervention and control groups, respectively.

Table 3. Secondary outcome measures at the 6- to 10-week follow-up.^a

Characteristics	Intervention		Usual care		<i>P</i> value for difference crude	<i>P</i> value for difference adjusted ^b
	Values	Change from baseline	Values	Change from baseline		
Risk factors						
SBP ^c (mm Hg), mean (SD)	123.6 (14.8)	-25.3 (27.4)	127.1 (13.3)	-16.5 (27.4)	.08	.02
DBP ^d (mm Hg), mean (SD)	75.3 (10.0)	-13.4 (15.6)	77.8 (8.7)	-9.1 (13.4)	.11	.05
BMI (kg/m ²), median (IQR)	26.3 (24.0 to 29.0)	-0.3 (-1.1 to 0.3)	27.0 (24.3 to 29.9)	-0.2 (-1.3 to 0.3)	.77	.39 ^e
Total cholesterol (mmol/L), mean (SD)	3.3 (0.8)	-1.5 (1.1)	3.3 (0.6)	-1.6 (0.9)	.48	.37
LDL ^f -cholesterol (mmol/L), mean (SD)	1.5 (0.6)	-1.4 (0.9)	1.5 (0.4)	-1.5 (0.8)	.28	.24
HDL ^g -cholesterol (mmol/L), median (IQR)	1.1 (0.9 to 1.5)	0.1 (-0.1 to 0.2)	1.2 (1.0 to 1.5)	0.1 (-0.1 to 0.3)	.80	.93
Triglycerides (mmol/L), median (IQR)	1.0 (0.8 to 1.5)	-0.3 (-0.8 to 0.0)	1.1 (0.8 to 1.4)	-0.3 (-0.7 to 0.0)	.93	.99
Fasting plasma glucose (mmol/L), median (IQR)	6.1 (5.8 to 6.8)	-1.4 (-3.3 to -0.3)	6.2 (5.6 to 7.0)	-0.9 (-2.8 to -0.2)	.81	.47
Self-reported parameters, mean (SD)						
Vegetable consumption	2.2 (0.7)	0.3 (0.8)	2.2 (0.7)	0.4 (1.1)	.74	.65
Fruit consumption	2.3 (0.7)	0.6 (1.0)	2.3 (0.7)	0.5 (1.1)	.29	.36
Fish consumption	2.1 (1.0)	0.9 (1.1)	2.0 (0.9)	0.8 (1.1)	.40	.18
Consumption of sweets	2.3 (0.9)	0.3 (0.9)	2.3 (0.9)	0.5 (1.1)	.27	.22
Healthy diet index	8.9 (2.0)	2.2 (2.3)	8.8 (1.7)	2.1 (2.6)	.82	.77

^aNumbers are presented as mean differences (SD), median differences (IQR), and *P* values. Crude and adjusted *P* values are shown for the differences in mean or median change (δ) between baseline and 6- to 10-week follow-up, comparing the intervention and usual care groups.

^bAdjusted for gender, age, weight, previous heart disease, and smoking status at the time of the index event.

^cSBP: systolic blood pressure.

^dDBP: diastolic blood pressure.

^eNot adjusted for weight.

^fLDL: low-density lipoprotein.

^gHDL: high-density lipoprotein.

Table 4. Secondary outcome measures at 12- to 14-month follow-up.^a

Characteristics	Intervention		Usual care		<i>P</i> value for difference crude	<i>P</i> value for difference adjusted ^b
	Values	Change from baseline	Values	Change from baseline		
Risk factors						
SBP ^c (mm Hg), mean (SD)	126.9 (16.0)	-24.0 (31.1)	126.5 (13.7)	-17.0 (28.3)	.22	.09
DBP ^d (mm Hg), mean (SD)	76.1 (10.9)	-12.6 (17.6)	75.7 (9.2)	-11.3 (17.1)	.69	.49
BMI (kg/m ²), median (IQR)	26.4 (23.6 to 30.1)	-0.3 (-1.3 to 1.0)	27.1 (24.8 to 29.1)	-0.5 (-1.2 to 0.8)	.57	.35 ^e
Total cholesterol (mmol/L), mean (SD)	3.5 (0.9)	-1.3 (1.2)	3.5 (0.8)	-1.5 (1.1)	.41	.30
LDL ^f -cholesterol (mmol/L), mean (SD)	2.1 (1.0)	-1.2 (1.1)	2.0 (0.9)	-1.4 (1.0)	.20	.21
HDL ^g -cholesterol (mmol/L), median (IQR)	1.2 (1.0 to 1.5)	0.1 (0.0 to 0.3)	1.3 (1.0 to 1.5)	0.1 (0.0 to 0.3)	.89	.77
Triglycerides (mmol/L), median (IQR)	1.0 (0.7 to 1.4)	-0.3 (-0.9 to 0.0)	1.2 (0.8 to 1.5)	-0.3 (-0.8 to 0.8)	.48	.98
Fasting plasma glucose (mmol/L), median (IQR)	6.0 (5.6 to 6.7)	-1.5 (-3.3 to -0.2)	6.0 (5.5 to 6.5)	-1.4 (-2.7 to 2.0)	.45	.48
HbA _{1c} ^h (mmol/mol), median (IQR)	40.0 (36.0 to 43.0)	1.0 (-0.2 to 3.0)	38.0 (37.0 to 43.0)	2.0 (-2.0 to 3.8)	.68	.40
Self-reported parameters, mean (SD)						
Vegetable consumption	2.1 (0.8)	0.3 (1.0)	2.0 (0.8)	0.2 (0.7)	.67	.77
Fruit consumption	2.0 (0.8)	0.3 (0.8)	2.2 (0.8)	0.3 (0.9)	.92	.99
Fish consumption	1.8 (1.1)	0.6 (1.0)	2.0 (0.9)	0.7 (1.0)	.70	.66
Consumption of sweets	2.3 (0.8)	0.4 (1.0)	2.1 (0.9)	0.3 (1.1)	.45	.37
Healthy diet index	8.2 (2.1)	1.6 (2.2)	8.2 (1.8)	1.4 (2.2)	.69	.45

^aNumbers are presented as mean differences (SD), median differences (q1, q3), and *P* values. Crude and adjusted *P* values are shown for the differences in mean or median change (δ) between baseline and 12- to 14-month follow-up, comparing the intervention and usual care groups.

^bAdjusted for gender, age, weight, previous heart disease, and smoking status at the time of the index event.

^cSBP: systolic blood pressure.

^dDBP: diastolic blood pressure.

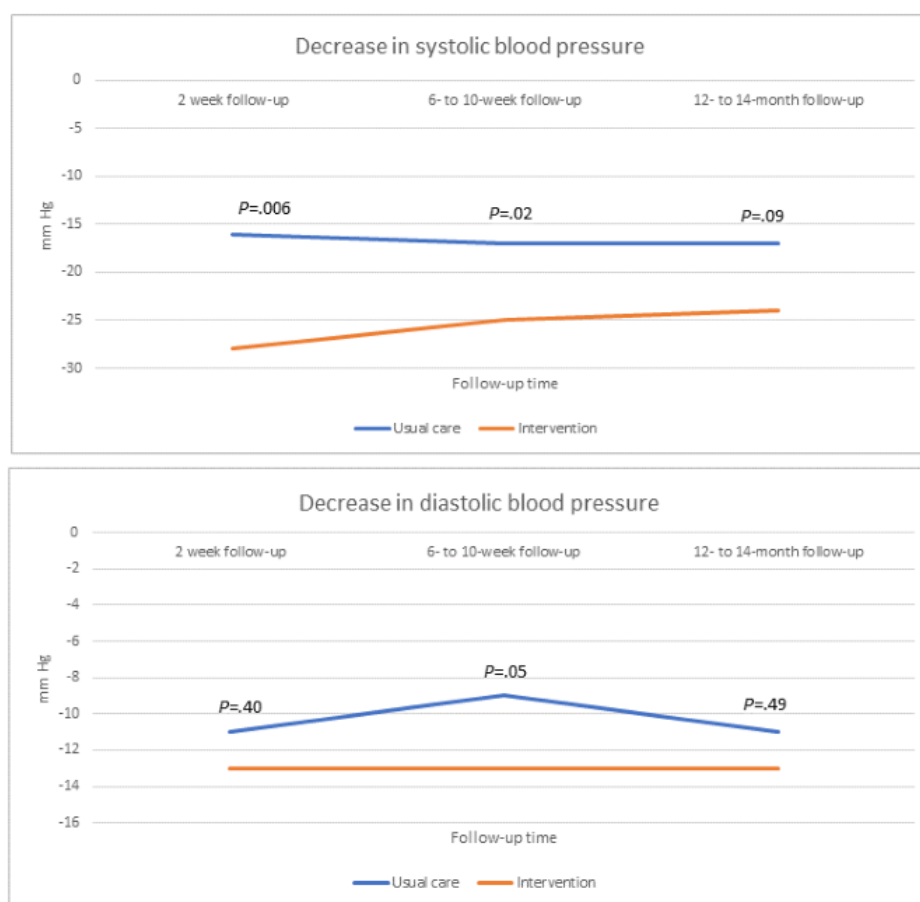
^eNot adjusted for weight.

^fLDL: low-density lipoprotein.

^gHDL: high-density lipoprotein.

^hHbA_{1c}: glycated hemoglobin A_{1c}.

Figure 5. Decrease in systolic and diastolic blood pressure between baseline and follow-up visits (absolute values). *P* values are adjusted for gender, age, weight, previous coronary heart disease, and smoking status at the time of the index event.

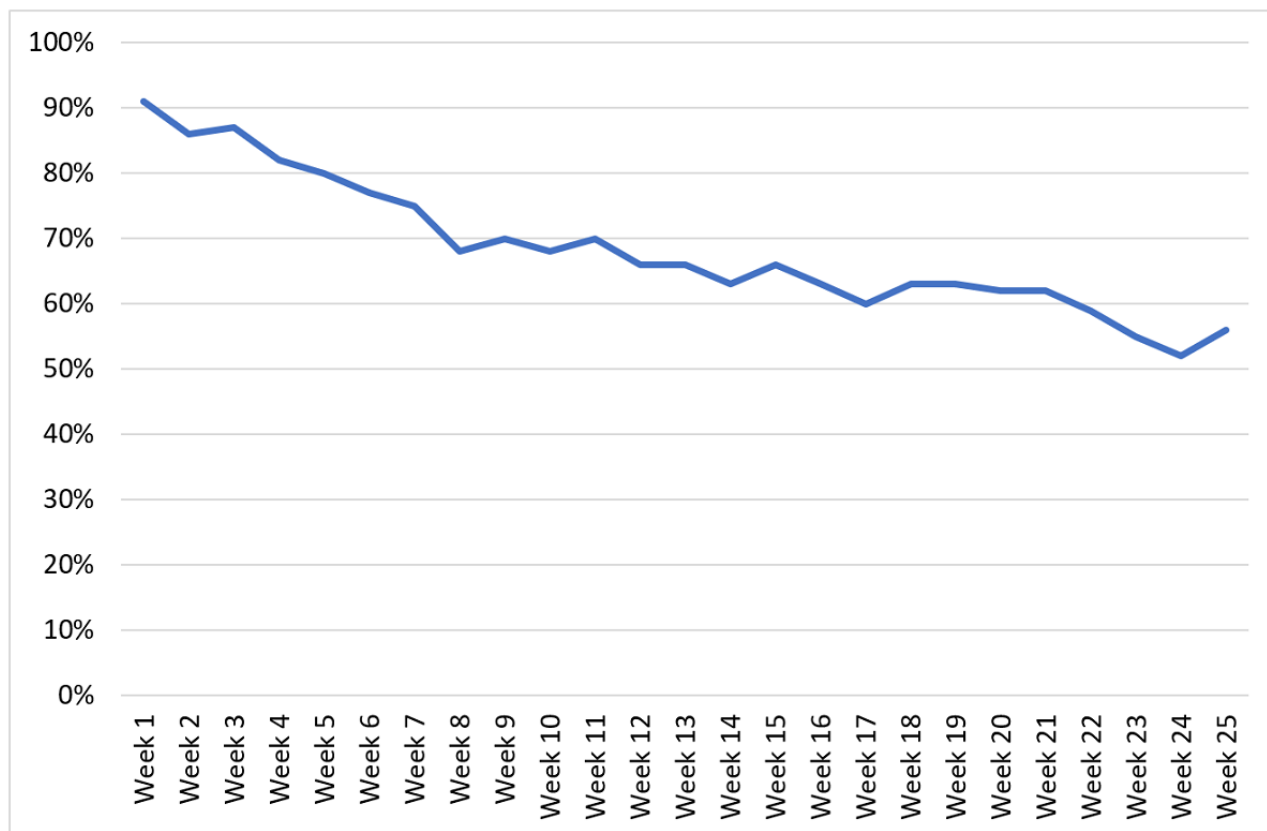


Uptake and Adherence

Attendance to the cardiac rehabilitation program was generally high. A total of 93% (93/100) of patients in the intervention group attended the first follow-up visit with a physiotherapist, and 71% (71/100) attended the second. The corresponding percentages for the usual care group were 98% (48/49) and 86% (42/49). However, only 66% (66/100) of patients in the intervention group and 82% (40/49) of patients in the usual care group performed a submaximal exercise test at both visits.

A total of 92% (92/100) of patients in the intervention group attended the 6- to 10-week follow-up and 96% (96/100) attended the 12- to 14-month follow-up, and the corresponding numbers for the usual care group were 98% (48/49) for both follow-up visits.

Uptake to the web application was 86.1% (87/101). Adherence (the proportion of the 87 patients who continued to log data at least twice per week) declined during the trial period, from its highest of 91% (79/87) at week 1 to 56% (49/87) at week 25 (Figure 6).

Figure 6. Percentage of patients logging data to the web-based application at least twice weekly.

The most frequent parameters to be logged to the web-based application by patients in the intervention group were intake of medication and consumption of vegetables, followed by physical activity ([Multimedia Appendix 1](#)).

Additional Analysis

There was no difference in the frequency of deaths due to any cause ($n=1$ in each group, $P=.62$). The frequencies of rehospitalizations during the 12- to 14-month follow-up period were also similar among the groups. In total, 22% (22/100) of the patients in the intervention group and 27% (13/49) of the patients in the usual care group were rehospitalized due to any cause ($P=.58$), and 8% (8/100) versus 4% (2/49) of the patients were hospitalized due to ischemic heart disease ($P=.35$).

Discussion

Principal Findings

In our study, evaluating the effects of a web-based application as an addition to usual care cardiac rehabilitation as compared with usual care alone, we found that there was a trend toward better exercise capacity in the intervention group; however, this was not statistically significant. On the other hand, the patients receiving access to the web-based application had a significantly reduced SBP at the first 2 follow-up visits, a reduced DBP at the second follow-up visit, and an initial increase in intake of fish and fruit. Although cardiac rehabilitation attendance was high in both groups, adherence to the web-based application declined over time.

After the baseline exercise test, all patients in our study received individualized physical activity recommendations and an individualized exercise prescription and were invited to participate in exercise-based cardiac rehabilitation based at the cardiac rehabilitation center. Those in the intervention group were, in addition, advised to rate their perceived exertion and time spent in exercise and log the information on the web-based application, after which they would receive automated positive feedback. Although attendance to physiotherapy follow-up visits was high, not all patients completed the submaximal exercise tests. The reasons for noncompletion were not officially documented, but according to clinical experience, reasons often include common cold, leg or joint pain, and perceived inability to exercise. During follow-up, both groups had a significant improvement in submaximal exercise capacity and, although not significant, there was a 4.1 W difference among the groups in favor of the intervention group. Increased physical activity levels using telehealth in cardiac rehabilitation have been demonstrated in previous studies. A recent meta-analysis showed that this applied especially to studies where telehealth was used as an adjunct to comprehensive cardiac rehabilitation [16]. The web-based application in our study did not specifically provide feedback on adherence to the exercise-based cardiac rehabilitation program but rather to increase physical activity levels in general, which may perhaps partly explain the lack of significant differences in exercise capacity among the groups.

The benefit of the combination of involvement of health care professionals with eHealth has been shown to be beneficial in studies addressing cardiovascular risk factors [31-33]. For example, a recent RCT found that participants using an

exercise-focused smartphone app had better fitness levels (measured by maximal oxygen consumption or VO_{2peak}) compared with usual care after a follow-up at the 1-year postcardiac rehabilitation [32]. The information that participants logged into the smartphone app was monitored by a physiotherapist who provided personal feedback. Adherence to the smartphone app was high, and participants felt that they reported to an individual involved in their care rather than a database or a robot. This knowledge can be used in the design of future studies aiming to increase the exercise-based cardiac rehabilitation adherence with the overall aim of increasing patients' physical fitness.

Most patients who had an MI were already on a regime of cardioprotective medication when discharged from the hospitals (Table 1). Despite this, a significant difference in BP control was observed among the groups, in favor of the intervention group. In our crude analysis, patients in the intervention group had a significantly larger SBP decrease between baseline and the 2-week follow-up. A numerical but nonsignificant difference among the groups then remained throughout the trial period. After adjusting for relevant covariates, the observed differences in BP control increased, in favor of the intervention group, and included a significant difference in SBP and DBP decrease between baseline and the 6- to 10-week follow-up. Patients in the intervention group achieved an 11.3 mm Hg larger decrease in SBP between baseline and the 2-week follow-up compared with patients in the usual care group, the difference being 8.8 mm Hg at the 6 to 10-week follow-up and 7.0 mm Hg at the 12- to 14-month follow-up. The choice of covariates was based on observed differences in baseline characteristics, which indicated a higher comorbidity in the intervention group. As such, not adjusting for baseline differences could have masked a difference among the groups. Hypertension is a predominant risk factor for CHD morbidity and mortality [7]. Previous meta-analyses have shown that a 10 mm Hg reduction in SBP reduces the risk of future CHD events by 22%, irrespective of the method of BP reduction [34]. Given the magnitude of the observed decrease in BP in our study, the difference is likely to have some clinical benefit. An improvement in reaching BP goals when having access to a web-based application as an aid in cardiac rehabilitation is in line with previous studies [11-13,20,35]. For example, Widmer et al [35] also demonstrated a substantial BP difference, with patients who after a percutaneous coronary intervention received digital health intervention (web-based or smartphone-based) as a complement to traditional cardiac rehabilitation having significant improvements in SBP (-10.8 mm Hg, $P<.001$ vs -6.1 mm Hg, $P=.36$) at a 3-month follow-up. Because intake of medication was the most frequently logged parameter in the web-based application, one possible reason for improved BP values in the intervention group might be an increased compliance to antihypertensive medication. Increased adherence to medication has been demonstrated in a previous study on a web-based application in cardiac rehabilitation [36]. In addition, increased patient engagement with self-monitoring has been shown to improve BP control [37,38]. The difference in BP among the groups declined throughout the trial period. One reason might be that adherence to behavioral recommendations has been seen

to decline over time in cardiac rehabilitation, or it could have something to do with decreasing adherence to the web-based application which is a known methodological challenge within eHealth [11,39]. As there was no measurement done at the end of the 25-week intervention period, we do not know exactly when the differences attenuated.

We also observed a significant beneficial effect on healthy dietary choices. During the first 2 weeks of follow-up, patients in the intervention group had a significantly higher score on the health diet index and higher intake of fish and fruit. Following a healthy diet rich in healthy oils and plant-based food has been shown to reduce cardiovascular events [40]. Previous studies on eHealth in cardiac rehabilitation have shown both improvement and no effect on dietary habits [15,16,41]. Here, the data were self-reported, which has considerable limitations, and the results should be interpreted with caution [42].

Smoking is a leading cause of preventable death globally and is a strong risk factor for CHD [7]. There was a numerical difference in the number of smokers who reported being abstinent at the follow-up visits, but owing to the low number of smokers in our study population, the differences were not statistically significant. In addition, whether the numerical difference was due to the intervention or chance finding cannot be stated. However, telehealth interventions in cardiac rehabilitation have been shown to reduce the likelihood of smoking by 23%, giving promise to using eHealth as a tool to tackle this important risk factor [16].

Adherence to eHealth

eHealth has the potential to improve participation and adherence to cardiac rehabilitation programs by including patients more actively in their own care and increasing flexibility and accessibility [21]. In our study, although uptake to the web-based application was high, adherence declined over time, with just over half of the users logging data at least twice weekly at the end of the 25-week intervention period. The main reason for limiting the intervention time to 25 weeks was to harmonize the intervention length with usual care, as most usual care cardiac rehabilitation interventions take place during the first 6 months after MI (nurse and physiotherapist visits, exercise training sessions, and patient education). In addition, patients usually have the highest level of motivation for lifestyle changes during the first months after MI [39]. The last reason was that adherence to eHealth interventions in cardiac rehabilitation has been shown to attenuate over time (weeks to months) [11,43]. In our study, reported reasons for stopping the use of the web-based application were mostly related to stress, some experienced a lack of feedback and some experienced too much feedback. User attrition in studies on eHealth is a well-known methodological challenge, with attrition rates reaching up to 60% to 80% [43]. High attrition rates can make it difficult to measure an intervention's effect and subsequently threaten both internal and external study validity [44,45]. Van der Mispel et al [44] studied user and website characteristics related to attrition. They demonstrated that attrition was higher for men and younger adults, as well as for less interactive components of the studied application. Buys et al [46] reported a general interest in technology-enabled home-based cardiac rehabilitation

among patients with cardiovascular disease. Their study demonstrated that patients with different characteristics were interested in different types of technology-based cardiac rehabilitation; for example, older patients were more interested in web-based options, and younger patients were more interested in app-based cardiac rehabilitation. They also looked at which parts of cardiac rehabilitation patients wanted to have technology-based options, including ideas on exercise, healthy meals, and stress management. This indicates that even if technology-based cardiac rehabilitation aims to make traditional programs more flexible and individualized, even the technology platforms need to be adjusted to individual needs for optimum effect.

Strengths and Limitations

In our study, we used data collected through standardized protocols from the SWEDEHEART registry [9,47]. The registry is well established in Swedish cardiac rehabilitation centers and is used daily by personnel. At the time of the study, SWEDEHEART nationwide coverage was >75%, registering eligible cases from 97% of Swedish hospitals. Data quality is regularly monitored, showing >95% agreement with data from hospital records [48]. Using SWEDEHEART prespecified time points for follow-up and procedures provided a standardization, which otherwise can be a challenge in multicenter trials. Cardiac

rehabilitation program attendance was high, and few participants were lost on follow-up.

The study was unblinded. Although blinding is preferred in RCTs, it is difficult to blind patients and health care providers in eHealth interventions. The follow-up data do not include information on the use of commercial web-based applications by the patients in the usual care group. The age limit considered in this study may restrict generalizability to all age groups. There might have been some selection bias, where more motivated patients would be more likely to agree to participate in a study using digital technology, requiring active participation. When interpreting the results, one should keep in mind the potential sources of inaccuracy in point BP measurements, which is a known source of bias in the clinical setting [49].

Conclusions

Digital technologies provide new opportunities in health care. Our results add to existing evidence and suggest that complementing comprehensive cardiac rehabilitation programs with a web-based application may positively affect risk factor outcomes and lifestyle, including BP and dietary choices. In addition, web-based technologies can be used to make cardiac rehabilitation programs more flexible and individualized, but further efforts should be invested to find ways to improve patient adherence to the platforms.

Acknowledgments

The study was funded by the Faculty of Medicine at Lund University, The Department of Coronary Care at Skåne University Hospital *Centre of Excellence in Skåne* Funds, and Avtal om Läkarutbildning och Forskning Grants County Council of Västerbotten. This study was partially sponsored by Cross Technology Solutions AB. The funding party had no role in design of the study, analysis of data, or writing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Percentage of reports per parameter in the web-based application.

[DOCX File, 16 KB - [jmir_v24i3e25224_app1.docx](#)]

Multimedia Appendix 2

CONSORT-eHEALTH checklist (V 1.6.1).

[PDF File (Adobe PDF File), 1222 KB - [jmir_v24i3e25224_app2.pdf](#)]

References

1. Nowbar AN, Gitto M, Howard JP, Francis DP, Al-Lamee R. Mortality from ischemic heart disease. *Circ Cardiovas Qual Outcomes* 2019 Jun;12(6):e005375. [doi: [10.1161/circoutcomes.118.005375](#)]
2. Townsend N, Nichols M, Scarborough P, Rayner M. Cardiovascular disease in Europe--epidemiological update 2015. *Eur Heart J* 2015 Oct 21;36(40):2696-2705. [doi: [10.1093/eurheartj/ehv428](#)] [Medline: [26306399](#)]
3. Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, et al. Explaining the decrease in U.S. deaths from coronary disease, 1980–2000. *N Engl J Med* 2007 Jun 07;356(23):2388-2398. [doi: [10.1056/nejmsa053935](#)]
4. Ambrosetti M, Abreu A, Corrà U, Davos C, Hansen D, Frederix I, et al. Secondary prevention through comprehensive cardiovascular rehabilitation: from knowledge to implementation. 2020 update. A position paper from the Secondary Prevention and Rehabilitation Section of the European Association of Preventive Cardiology. *Eur J Prev Cardiol* 2020 Apr 07:2020. [doi: [10.1177/2047487320913379](#)] [Medline: [33611446](#)]

5. Wood D, Kotseva K, Connolly S, Jennings C, Mead A, Jones J, et al. Nurse-coordinated multidisciplinary, family-based cardiovascular disease prevention programme (EUROACTION) for patients with coronary heart disease and asymptomatic individuals at high risk of cardiovascular disease: a paired, cluster-randomised controlled trial. *Lancet* 2008 Jun;371(9629):1999-2012. [doi: [10.1016/s0140-6736\(08\)60868-5](https://doi.org/10.1016/s0140-6736(08)60868-5)]
6. Piepoli MF, Corrà U, Adamopoulos S, Benzer W, Bjarnason-Wehrens B, Cupples M, et al. Secondary prevention in the clinical management of patients with cardiovascular diseases. Core components, standards and outcome measures for referral and delivery: a policy statement from the cardiac rehabilitation section of the European Association for Cardiovascular Prevention and Rehabilitation. Endorsed by the Committee for Practice Guidelines of the European Society of Cardiology. *Eur J Prev Cardiol* 2014 Jun 20;21(6):664-681. [doi: [10.1177/2047487312449597](https://doi.org/10.1177/2047487312449597)] [Medline: [22718797](https://pubmed.ncbi.nlm.nih.gov/22718797/)]
7. Piepoli MF. 2016 European guidelines on cardiovascular disease prevention in clinical practice: the sixth joint task force of the European society of cardiology and other societies on cardiovascular disease prevention in clinical practice (constituted by representatives of 10 societies and by invited experts). *Int J Behav Med* 2017 Jun 13;24(3):321-419. [doi: [10.1007/s12529-016-9583-6](https://doi.org/10.1007/s12529-016-9583-6)] [Medline: [28290077](https://pubmed.ncbi.nlm.nih.gov/28290077/)]
8. Kotseva K, Wood D, De Bacquer D, De Backer G, Rydén L, Jennings C, EUROASPIRE Investigators. EUROASPIRE IV: a European Society of Cardiology survey on the lifestyle, risk factor and therapeutic management of coronary patients from 24 European countries. *Eur J Prev Cardiol* 2016 Apr;23(6):636-648. [doi: [10.1177/2047487315569401](https://doi.org/10.1177/2047487315569401)] [Medline: [25687109](https://pubmed.ncbi.nlm.nih.gov/25687109/)]
9. Jernberg T, Boberg B, Bäck M, Dahlbom L, Erlinge D, Friberg O. SWEDEHEART Årsrapport 2019. Uppsala Clinical Research Center, UCR. 2020. URL: https://www-ucr-uu-se.translate.goog/swedeheart/arsrapport?_x_tr_sl=sv&_x_tr_tl=en&_x_tr_hl=en&_x_tr_pto=sc [accessed 2022-02-11]
10. Michelsen HO, Sjölin I, Schlyter M, Hagström E, Kiessling A, Henriksson P, et al. Cardiac rehabilitation after acute myocardial infarction in Sweden - evaluation of programme characteristics and adherence to European guidelines: the Perfect Cardiac Rehabilitation (Perfect-CR) study. *Eur J Prev Cardiol* 2020 Jan 26;27(1):18-27. [doi: [10.1177/2047487319865729](https://doi.org/10.1177/2047487319865729)] [Medline: [31349776](https://pubmed.ncbi.nlm.nih.gov/31349776/)]
11. Blasco A, Carmona M, Fernández-Lozano I, Salvador C, Pascual M, Sagredo P, et al. Evaluation of a telemedicine service for the secondary prevention of coronary artery disease. *J Cardiopulm Rehabil Prev* 2012;32(1):25-31. [doi: [10.1097/HCR.0b013e3182343aa7](https://doi.org/10.1097/HCR.0b013e3182343aa7)] [Medline: [22113368](https://pubmed.ncbi.nlm.nih.gov/22113368/)]
12. Pietrzak E, Cotea C, Pullman S. Primary and secondary prevention of cardiovascular disease: is there a place for Internet-based interventions? *J Cardiopulm Rehabil Prev* 2014;34(5):303-317. [doi: [10.1097/HCR.0000000000000063](https://doi.org/10.1097/HCR.0000000000000063)] [Medline: [25079147](https://pubmed.ncbi.nlm.nih.gov/25079147/)]
13. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. *Heart* 2014 Nov 27;100(22):1770-1779 [FREE Full text] [doi: [10.1136/heartjnl-2014-305783](https://doi.org/10.1136/heartjnl-2014-305783)] [Medline: [24973083](https://pubmed.ncbi.nlm.nih.gov/24973083/)]
14. Neubeck L, Redfern JU, Fernandez R, Briffa T, Bauman A, Freedman SB. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review. *Eur J Cardiovas Preven Rehabil* 2009 Jun;16(3):281-289. [doi: [10.1097/hjr.0b013e32832a4e7a](https://doi.org/10.1097/hjr.0b013e32832a4e7a)]
15. Brørs G, Pettersen TR, Hansen TB, Fridlund B, Hølvold LB, Lund H, et al. Modes of e-Health delivery in secondary prevention programmes for patients with coronary artery disease: a systematic review. *BMC Health Serv Res* 2019 Jun 10;19(1):364 [FREE Full text] [doi: [10.1186/s12913-019-4106-1](https://doi.org/10.1186/s12913-019-4106-1)] [Medline: [31182100](https://pubmed.ncbi.nlm.nih.gov/31182100/)]
16. Jin K, Khonsari S, Gallagher R, Gallagher P, Clark AM, Freedman B, et al. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review and meta-analysis. *Eur J Cardiovas Nurs* 2019 Apr 22;18(4):260-271. [doi: [10.1177/1474515119826510](https://doi.org/10.1177/1474515119826510)] [Medline: [30667278](https://pubmed.ncbi.nlm.nih.gov/30667278/)]
17. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, et al. Current science on consumer use of mobile health for cardiovascular disease prevention. *Circulation* 2015 Sep 22;132(12):1157-1213. [doi: [10.1161/cir.0000000000000232](https://doi.org/10.1161/cir.0000000000000232)]
18. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. *J Am Heart Assoc* 2013 Nov 18;2(6):e000568. [doi: [10.1161/jaha.113.000568](https://doi.org/10.1161/jaha.113.000568)]
19. Widmer RJ, Senecal C, Allison TG, Lopez-Jimenez F, Lerman LO, Lerman A. Dose-response effect of a digital health intervention during cardiac rehabilitation: subanalysis of randomized controlled trial. *J Med Internet Res* 2020 Feb 26;22(2):e13055 [FREE Full text] [doi: [10.2196/13055](https://doi.org/10.2196/13055)] [Medline: [32130116](https://pubmed.ncbi.nlm.nih.gov/32130116/)]
20. 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](https://doi.org/10.2196/16695)] [Medline: [32628615](https://pubmed.ncbi.nlm.nih.gov/32628615/)]
21. Frederix I, Caiani EG, Dendale P, Anker S, Bax J, Böhm A, et al. ESC e-Cardiology Working Group Position Paper: overcoming challenges in digital health implementation in cardiovascular medicine. *Eur J Prev Cardiol* 2019 Jul 27;26(11):1166-1177. [doi: [10.1177/2047487319832394](https://doi.org/10.1177/2047487319832394)] [Medline: [30917695](https://pubmed.ncbi.nlm.nih.gov/30917695/)]
22. Gonzalez M, Sjölin I, Bäck M, Michelsen HO, Tanha T, Sandberg C, et al. Effect of a lifestyle-focused electronic patient support application for improving risk factor management, self-rated health, and prognosis in post-myocardial infarction patients: study protocol for a multi-center randomized controlled trial. *Trials* 2019 Jan 24;20(1):76 [FREE Full text] [doi: [10.1186/s13063-018-3118-1](https://doi.org/10.1186/s13063-018-3118-1)] [Medline: [30678709](https://pubmed.ncbi.nlm.nih.gov/30678709/)]

23. De Schutter A, Kachur S, Lavie C, Menezes A, Shum K, Bangalore S, et al. Cardiac rehabilitation fitness changes and subsequent survival. *Eur Heart J Qual Care Clin Outcomes* 2018 Jul 01;4(3):173-179. [doi: [10.1093/ehjqcco/qcy018](https://doi.org/10.1093/ehjqcco/qcy018)] [Medline: [29701805](https://pubmed.ncbi.nlm.nih.gov/29701805/)]
24. SWEDEHEART Annual Report 2020. Uppsala Clinical Research Cente. 2020. URL: <https://www.ucr.uu.se/swedeheart/> [accessed 2022-02-11]
25. No authors listed. Exercise tests in relation to cardiovascular function. *Ann Intern Med* 1968 Sep 01;69(3):652. [doi: [10.7326/0003-4819-69-3-652_2](https://doi.org/10.7326/0003-4819-69-3-652_2)]
26. Michelsen S. Reproducibility of cumulative work, heart rate and blood pressure response during stepwise versus continuous load increment during a maximal bicycle ergometer test. *Scand J Clin Lab Invest* 1990 Jun 08;50(4):409-415. [doi: [10.3109/00365519009091599](https://doi.org/10.3109/00365519009091599)] [Medline: [2392654](https://pubmed.ncbi.nlm.nih.gov/2392654/)]
27. Borg G. Borg's Perceived Exertion and Pain Scales. Champaign, Illinois: Human Kinetics; 1998.
28. Nationella Riktlinjer för Prevention Och Behandling Vid Ohälsosamma Levnadsvanor - Stöd för Styrning Och Ledning. Stockholm: Socialstyrelsen; 2018:470-491.
29. Haskell WL, Lee IM, Pate RR, Powell KE, Blair SN, Franklin BA, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. *Circulation* 2007 Aug 28;116(9):1081-1093. [doi: [10.1161/circulationaha.107.185649](https://doi.org/10.1161/circulationaha.107.185649)]
30. Grimby G, Frändin K. On the use of a six-level scale for physical activity. *Scand J Med Sci Sports* 2018 Mar 17;28(3):819-825. [doi: [10.1111/sms.12991](https://doi.org/10.1111/sms.12991)] [Medline: [29027263](https://pubmed.ncbi.nlm.nih.gov/29027263/)]
31. Green BB, Cook AJ, Ralston JD, Fishman PA, Catz SL, Carlson J, et al. Effectiveness of home blood pressure monitoring, web communication, and pharmacist care on hypertension control: a randomized controlled trial. *J Am Med Assoc* 2008 Jun 25;299(24):2857-2867 [FREE Full text] [doi: [10.1001/jama.299.24.2857](https://doi.org/10.1001/jama.299.24.2857)] [Medline: [18577730](https://pubmed.ncbi.nlm.nih.gov/18577730/)]
32. Lunde P, Bye A, Bergland A, Grimsmo J, Jarstad E, Nilsson BB. Long-term follow-up with a smartphone application improves exercise capacity post cardiac rehabilitation: a randomized controlled trial. *Eur J Prev Cardiol* 2020 Nov 28;27(16):1782-1792 [FREE Full text] [doi: [10.1177/2047487320905717](https://doi.org/10.1177/2047487320905717)] [Medline: [32106713](https://pubmed.ncbi.nlm.nih.gov/32106713/)]
33. Widmer RJ, Allison TG, Lennon R, Lopez-Jimenez F, Lerman LO, Lerman A. Digital health intervention during cardiac rehabilitation: a randomized controlled trial. *Am Heart J* 2017 Jun;188:65-72. [doi: [10.1016/j.ahj.2017.02.016](https://doi.org/10.1016/j.ahj.2017.02.016)] [Medline: [28577682](https://pubmed.ncbi.nlm.nih.gov/28577682/)]
34. Law MR, Morris JK, Wald NJ. Use of blood pressure lowering drugs in the prevention of cardiovascular disease: meta-analysis of 147 randomised trials in the context of expectations from prospective epidemiological studies. *Br Med J* 2009 May 19;338(may19 1):b1665 [FREE Full text] [doi: [10.1136/bmj.b1665](https://doi.org/10.1136/bmj.b1665)] [Medline: [19454737](https://pubmed.ncbi.nlm.nih.gov/19454737/)]
35. Widmer RJ, Allison TG, Lerman LO, Lerman A. Digital health intervention as an adjunct to cardiac rehabilitation reduces cardiovascular risk factors and rehospitalizations. *J Cardiovasc Transl Res* 2015 Jul 7;8(5):283-292 [FREE Full text] [doi: [10.1007/s12265-015-9629-1](https://doi.org/10.1007/s12265-015-9629-1)] [Medline: [25946990](https://pubmed.ncbi.nlm.nih.gov/25946990/)]
36. 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](https://doi.org/10.1016/j.ahj.2016.05.005)] [Medline: [27502855](https://pubmed.ncbi.nlm.nih.gov/27502855/)]
37. McCartney DE, McManus RJ. Self-monitoring and self-management. *Curr Opin Nephrol Hypertens* 2016;25(6):502-507. [doi: [10.1097/mnh.0000000000000279](https://doi.org/10.1097/mnh.0000000000000279)]
38. Uhlig K, Patel K, Ip S, Kitsios GD, Balk EM. Self-measured blood pressure monitoring in the management of hypertension: a systematic review and meta-analysis. *Ann Intern Med* 2013 Aug 06;159(3):185-194 [FREE Full text] [doi: [10.7326/0003-4819-159-3-201308060-00008](https://doi.org/10.7326/0003-4819-159-3-201308060-00008)] [Medline: [23922064](https://pubmed.ncbi.nlm.nih.gov/23922064/)]
39. Andrikopoulos G, Tzeis S, Nikas N, Richter D, Pipilis A, Gotsis A, et al. Short-term outcome and attainment of secondary prevention goals in patients with acute coronary syndrome--results from the countrywide TARGET study. *Int J Cardiol* 2013 Sep 30;168(2):922-927. [doi: [10.1016/j.ijcard.2012.10.049](https://doi.org/10.1016/j.ijcard.2012.10.049)] [Medline: [23186596](https://pubmed.ncbi.nlm.nih.gov/23186596/)]
40. Estruch R, Ros E, Salas-Salvadó J, Covas M, Corella D, Arós F, et al. Primary prevention of cardiovascular disease with a Mediterranean diet. *N Engl J Med* 2013 Apr 04;368(14):1279-1290. [doi: [10.1056/nejmoa1200303](https://doi.org/10.1056/nejmoa1200303)]
41. Lear S, Singer J, Banner-Lukaris D, Horvat D, Park J, Bates J, et al. Improving access to cardiac rehabilitation using the internet: a randomized trial. *Stud Health Technol Inform* 2015;209:58-66. [Medline: [25980706](https://pubmed.ncbi.nlm.nih.gov/25980706/)]
42. Warren JM, Ekelund U, Besson H, Mezzani A, Geladas N, Vanhees L. Assessment of physical activity – a review of methodologies with reference to epidemiological research: a report of the exercise physiology section of the European Association of Cardiovascular Prevention and Rehabilitation. *Eur J Cardiovas Preven Rehabil* 2010 Apr;17(2):127-139. [doi: [10.1097/hjr.0b013e32832ed875](https://doi.org/10.1097/hjr.0b013e32832ed875)]
43. Geraghty AW, Torres LD, Leykin Y, Pérez-Stable EJ, Muñoz RF. Understanding attrition from international Internet health interventions: a step towards global eHealth. *Health Promot Int* 2013 Sep 10;28(3):442-452 [FREE Full text] [doi: [10.1093/heapro/das029](https://doi.org/10.1093/heapro/das029)] [Medline: [22786673](https://pubmed.ncbi.nlm.nih.gov/22786673/)]
44. Van der Mispel C, Poppe L, Crombez G, Verloigne M, De Bourdeaudhuij I. A self-regulation-based ehealth intervention to promote a healthy lifestyle: investigating user and website characteristics related to attrition. *J Med Internet Res* 2017 Jul 11;19(7):e241 [FREE Full text] [doi: [10.2196/jmir.7277](https://doi.org/10.2196/jmir.7277)] [Medline: [28698168](https://pubmed.ncbi.nlm.nih.gov/28698168/)]

45. Eysenbach G. The law of attrition. *J Med Internet Res* 2005 Mar 31;7(1):e11 [FREE Full text] [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](https://pubmed.ncbi.nlm.nih.gov/15829473/)]
46. Buys R, Claes J, Walsh D, Cornelis N, Moran K, Budts W, et al. Cardiac patients show high interest in technology enabled cardiovascular rehabilitation. *BMC Med Inform Decis Mak* 2016 Jul 19;16(1):95 [FREE Full text] [doi: [10.1186/s12911-016-0329-9](https://doi.org/10.1186/s12911-016-0329-9)] [Medline: [27431419](https://pubmed.ncbi.nlm.nih.gov/27431419/)]
47. Physiotherapy SEPHIA. Uppsala Clinical Research Center, Uppsala, Sweden. 2018. URL: <https://www.ucr.uu.se/swedeheart/dokument-sephia/fysioterapi-sephia> [accessed 2022-02-11]
48. SWEDEHEART. Uppsala Clinical Research Center, Uppsala. 2020. URL: <https://www.ucr.uu.se/swedeheart/> [accessed 2022-02-11]
49. Kallioinen N, Hill A, Horswill M, Ward H, Watson M. Sources of inaccuracy in the measurement of adult patients' resting blood pressure in clinical settings: a systematic review. *J Hypertens* 2017 Mar;35(3):421-441 [FREE Full text] [doi: [10.1097/HJH.0000000000001197](https://doi.org/10.1097/HJH.0000000000001197)] [Medline: [27977471](https://pubmed.ncbi.nlm.nih.gov/27977471/)]

Abbreviations

BP: blood pressure

CHD: coronary heart disease

DBP: diastolic blood pressure

LDL: low-density lipoprotein

MI: myocardial infarction

RCT: randomized controlled trial

SBP: systolic blood pressure

SPIRIT: Standard Protocol Items: Recommendations for Interventional Trials

SWEDEHEART: Swedish Web-System for Enhancement and Development of Evidence-Based Care in Heart Disease Evaluated According to Recommended Therapies

Edited by G Eysenbach; submitted 23.10.20; peer-reviewed by T Ratz, B Bente, D Fletcher; comments to author 08.01.21; revised version received 09.02.21; accepted 28.12.21; published 31.03.22.

Please cite as:

*Ögmundsdóttir Michelsen H, Sjölin I, Bäck M, Gonzalez Garcia M, Olsson A, Sandberg C, Schiopu A, Leósdóttir M
Effect of a Lifestyle-Focused Web-Based Application on Risk Factor Management in Patients Who Have Had a Myocardial Infarction:
Randomized Controlled Trial*

J Med Internet Res 2022;24(3):e25224

URL: <https://www.jmir.org/2022/3/e25224>

doi: [10.2196/25224](https://doi.org/10.2196/25224)

PMID: [35357316](https://pubmed.ncbi.nlm.nih.gov/35357316/)

©Halldóra Ögmundsdóttir Michelsen, Ingela Sjölin, Maria Bäck, Manuel Gonzalez Garcia, Anneli Olsson, Camilla Sandberg, Alexandru Schiopu, Margrét Leósdóttir. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 31.03.2022. 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.

Original Paper

The Characteristics and Functionalities of Mobile Apps Aimed at Patients Diagnosed With Immune-Mediated Inflammatory Diseases: Systematic App Search

Rosa Romero-Jimenez¹, PharmD, PhD; Vicente Escudero-Vilaplana¹, PharmD, PhD; Esther Chamorro-De-Vega¹, PharmD, PhD; Arantza Ais-Larisgoitia¹, PharmD; Maria Elena Lobato Matilla¹, PharmD; Ana Herranz-Alonso¹, PharmD, PhD; Maria Sanjurjo¹, PharmD

Instituto de Investigación Sanitaria Gregorio Marañón, Hospital General Universitario Gregorio Marañón, Madrid, Spain

Corresponding Author:

Rosa Romero-Jimenez, PharmD, PhD
Instituto de Investigación Sanitaria Gregorio Marañón
Hospital General Universitario Gregorio Marañón
Doctor Esquerdo, 46
Madrid, 28007
Spain
Phone: 34 915868414
Email: rosa.romero@salud.madrid.org

Abstract

Background: Immune-mediated inflammatory diseases (IMIDs) are systemic conditions associated with a high social and health impact. New treatments have changed the prognosis of IMIDs and have increased patient autonomy in disease management. Mobile apps have enormous potential to improve health outcomes in patients with IMIDs. Although a large number of IMID apps are available, the app market is not regulated, and functionality and reliability remain uncertain.

Objective: Our aims are to review available apps for patients with IMIDs or caregivers and to describe the main characteristics and functionalities of these apps.

Methods: We performed an observational, cross-sectional, descriptive study of all apps for patients with IMIDs. Between April 5 and 14, 2021, we conducted a search of the App Store (iOS) and Play Store (Android) platforms. We used the names of the different IMIDs as search terms. The inclusion criteria were as follows: content related to IMIDs, English or Spanish language, and user population consisting of patients and health care consumers, including family and caregivers. The variables analyzed were as follows: app name, type of IMID, platform (Android or iOS), country of origin, language, category of the app, cost, date of the last update, size, downloads, author affiliation, and functionalities.

Results: We identified 713 apps in the initial search, and 243 apps met the criteria and were analyzed. Of these, 37% (n=90) were on Android, 27.2% (n=66) on iOS, and 35.8% (n=87) on both platforms. The most frequent categories were health and well-being/fitness apps (n=188, 48.5%) and medicine (n=82, 37.9%). A total of 211 (82.3%) apps were free. The mean time between the date of the analysis and the date of the most recent update was 18.5 (SD 19.3) months. Health care professionals were involved in the development of 100 (41.1%) apps. We found differences between Android and iOS in the mean time since the last update (16.2, SD 14.7 months vs 30.3, SD 25.7 months) and free apps (85.6% vs 75.8%; respectively). The functionalities were as follows: general information about lifestyles, nutrition, or exercises (n=135, 55.6%); specific information about the disease or treatment (n=102, 42%); recording of symptoms or adverse events (n=51, 21%); agenda/calendar (n=44, 18.1%); reminder medication (n=41, 16.9%); and recording of patient-reported outcomes (n=41, 16.9%). A total of 147 (60.5%) apps had more than one functionality.

Conclusions: IMID-related apps are heterogeneous in terms of functionality and reliability. Apps may be a useful complement to IMID care, especially inpatient education (their most frequent functionality). However, more than half of the IMID apps had not been developed by health care professionals or updated in the last year.

(*J Med Internet Res* 2022;24(3):e31016) doi:[10.2196/31016](https://doi.org/10.2196/31016)

KEYWORDS

immune-mediated inflammatory disease; mobile app; mHealth; mobile health; chronic disease; disease management; outcomes; functionality; quality; patient education; health outcomes; reliability

Introduction

Immune-mediated inflammatory diseases (IMIDs) are systemic conditions characterized by altered immune regulation causing chronic inflammation in specific organs or systems [1]. IMIDs include rheumatologic diseases (eg, rheumatoid arthritis, ankylosing spondylitis, and psoriatic arthritis), digestive diseases (eg, Crohn disease and ulcerative colitis), and dermatologic diseases (eg, psoriasis). It is estimated that 5% to 7% of the world's population has an IMID [1]. Besides, IMIDs are associated with a high social and health impact in terms of morbidity and mortality, quality of life, and psychological and occupational aspects.

Management and treatment of IMIDs have changed substantially in the last decade, mainly due to the emergence of biological therapies [2]. However, since these new treatments are associated with problems related to administration, toxicity, and adherence, patients should receive adequate training and information. Besides, patients are becoming increasingly active and require more information. A survey conducted in the 27 countries of the European Union revealed that, in the previous 3 months to the survey, 53% of citizens sought online health information related to injury, disease, nutrition, improving health, and similar data [3].

Mobile health (mHealth) technologies incorporating strategies for remote self-management may offer an effective alternative to classic outpatient-based approaches [4]. In particular, broad accessibility to mobile apps enables them to complement health care 24 hours a day, 7 days a week at low cost [5-7]. Apps help to better understand and manage chronic diseases such as IMIDs [8]. Besides, patients with IMIDs are younger than patients with other chronic diseases, and they are more familiar with these technologies [1]. Thus, apps represent an opportunity to improve self-management of care and improve their quality of life.

Health care professionals could play an essential role not only in the review or verification of the contents of these apps but also in their prescription to the best candidates and the recommendation of the most reliable options. However, the lack of studies providing detailed IMID-related app characteristics limits health care professionals' knowledge in this field. Available IMID apps vary substantially in terms of features, functionality, and reliability [4,9-14]. Up-to-date information about the current situation of apps for patients with IMIDs can be a substantial help to health care professionals in guiding patients and identifying possible risks derived from their use, as well as identifying the needs and directions for future development of these tools. Thus, our objectives were to provide a review of the apps available in the marketplace for patients with IMIDs or caregivers and to identify and describe the main characteristics and functionalities of these apps.

Methods

We performed an observational, cross-sectional, descriptive study of all the smartphone apps designed for patients with IMIDs available on the iOS and Android platforms.

The methodology used for the selection of the apps followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) system. We searched the iOS Apple Store and Android Google Play Store from April 5 to 14, 2021 using the following terms: "ankylosing spondylitis," "Crohn's disease," "IBD," "inflammatory bowel disease," "immune-mediated inflammatory diseases," "immune-mediated inflammatory disorders," "psoriasis," "psoriatic arthritis," "rheumatoid arthritis," and "ulcerative colitis." As the resultant app list was potentially endless (similar to a Google search), we used the approach followed in previous reviews and performed the screening process until 20 consecutive apps yielded no new potentially relevant apps [15,16].

Two researchers with experience in the analysis of apps, design, and development, and in the management of patients with IMIDs (authors RRJ and VEV) performed identical searches independently. The inclusion criteria were as follows: content related to IMIDs, English or Spanish language, and user population consisting of patients and health consumers, including family or caregivers. Those apps that did not target patients (ie, apps targeting health professionals) were excluded.

The variables analyzed were the name of the app, type of IMID (ankylosing spondylitis, Crohn disease, psoriasis, psoriatic arthritis, rheumatoid arthritis, or ulcerative colitis), platform (Android or iOS), country of origin, language, category of the app, cost, date of the last update, size, downloads, author affiliation, and functionalities. The author's affiliation was classified as follows: university, health administration or government, hospital/clinic, foundation, patient association, pharmaceutical company, and technology company. Concerning technology companies, we identified two categories depending on whether they were health related or not. Functionalities were further classified into the following categories: information (disease or treatment and lifestyle, nutrition, and exercises), self-diagnosis, disease activity monitoring, symptom or patient-reported outcome (PRO) recording, medication reminders, adherence monitoring, agenda/calendar (a diary or calendar with or without function to record an appointment), contact with health care professionals, and social network. The app search method was developed by the authors and has not been validated. This methodology to analyze app characteristics has been used in other similar studies [4,16,17].

The data were collected from the description and the screenshots available at the app store websites. In cases where one of the variables was not available or there was any doubt, the app was downloaded onto an iPhone 11 (version 14.4.2) or a Xiaomi Redmi Note 7 (version 9.0). For the categorization of author

affiliation, we conducted Google searches using the name or the website provided by the app store.

The researchers (RRJ and VEV) evaluated each app independently. Data were analyzed using SPSS Statistics for Windows, Version 21.0 (IBM Corp; descriptive statistics). The homogeneity of the groups was analyzed using a univariate analysis by applying the chi-square test to compare qualitative variables and the *t* test or Mann-Whitney test to compare quantitative variables. A *P* value <.05 was considered statistically significant. Cohen kappa (κ) test was performed to determine the reliability of the data analyzed by these two independent researchers (RRJ and VEV).

The two researchers jointly analyzed those apps again in which the information collected varied between them. Discrepancies were discussed and an agreement was reached.

Results

General Characteristics

We identified 713 apps in the initial search in the two app stores; of these, 470 were excluded (Figure 1). We finally analyzed

243 apps: 37% (n=90) on Android, 27.2% (n=66) on iOS, and 35.8% (n=87) on both platforms. Concerning app characteristics, the agreement between the two researchers was excellent ($\kappa=0.868$).

Most of the apps belonged to the health and well-being/fitness and medicine categories. The mean time since the last update was 18.5 (SD 19.3) months, and the mean app size was 28.6 (SD 46.9) Mb. A total of 43 (17.7%) apps required payment for use, with a mean price of US \$10.30 (SD US \$12.00). Concerning the language, 86.8% (n=211) of the apps analyzed were in only one language, and 15 (6.2%) were in three or more languages. A total of 234 (96%) apps were in English, and 11.4% (29/243) were in Spanish. Table 1 shows the remaining general app characteristics.

We found statistically significant differences (*P*<.001) between the Android and iOS apps, as follows: time since the last update, 16.2 (SD 14.7) months vs 30.3 (SD 25.7) months; apps with cost, 14.4% vs. 24.2%; and size, 11.2 (SD 13.9) Mb vs 50.2 (SD 77.9) Mb.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram and app selection. IMID: immune-mediated inflammatory disease.

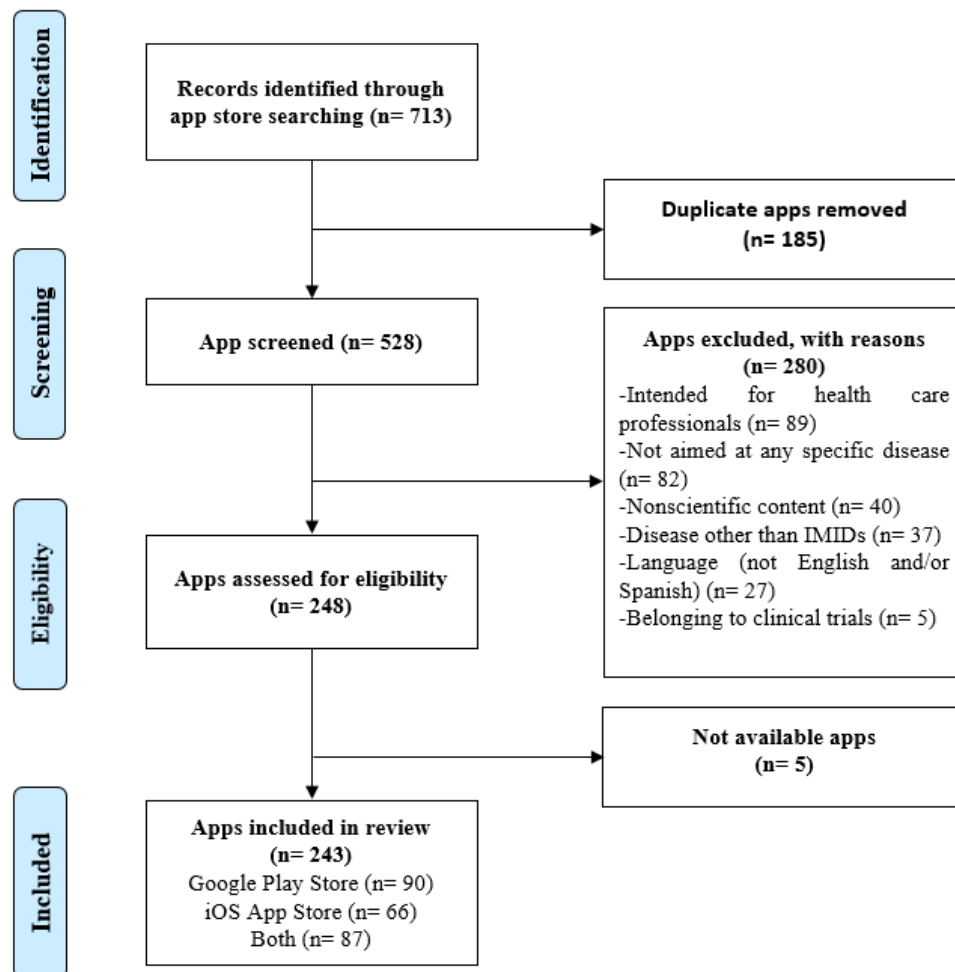


Table 1. App characteristics.

Characteristics	Apps (N=243), n (%)
Store category	
Health and well-being/fitness	118 (48.5)
Medicine	92 (37.9)
Food and drink	8 (3.3)
Books and references	7 (3.0)
Lifestyle	6 (2.5)
Social networking	4 (1.6)
Education	3 (1.2)
Other	5 (2.0)
Time since last update (months)	
<1	23 (9.5)
2-3	27 (11.1)
4-6	24 (9.9)
6-12	37 (15.2)
12-24	55 (22.6)
>24	61 (25.1)
Not available	16 (6.6)
Downloads	
0-100	25 (10.3)
100-500	32 (13.2)
500-1000	26 (10.7)
1000-5000	37 (15.2)
5000-10,000	19 (7.8)
10,000-50,000	23 (9.5)
50,000-100,000	2 (0.8)
100,000-500,000	10 (4.1)
>500,000	3 (1.2)
Not available	66 (27.2)
Cost	
No	200 (82.3)
Yes	43 (17.7)

App Author Affiliation

The most frequent type of author affiliation was a nonhealth technology company (n=77, 31.7%; [Table 2](#)). The author

affiliation was not available in 27.2% (n=66) of apps. Concerning the country of origin, 38.3% (n=93) of apps were from the United States, 8.2% (n=20) were from India, and 7.8% (n=19) were from the United Kingdom.

Table 2. App author affiliation.

Characteristics	Apps (N=243), n (%)
Type of author/developer	
Nonhealth technology company	77 (31.7)
Health-related technology company	57 (23.5)
Pharmaceutical company	12 (4.9)
Hospital/clinic	6 (2.5)
Foundation	18 (7.4)
Patient association	4 (1.6)
Health administration/government	2 (0.8)
University	1 (0.4)
Not available	66 (27.2)
Continent of origin	
North America	106 (43.6)
Europe	55 (22.6)
Asia	32 (13.2)
Oceania	5 (2.1)
South America	1 (0.4)
Not available	44 (18.1)

Type of IMIDs Targeted by the Apps

A total of 135 (55.5%) apps were specifically aimed at one IMID, and 21 (8.6%) were aimed at all IMIDs (Table 3).

Concerning author affiliation, we found statistically significant differences between specific ankylosing spondylitis-related apps and specific psoriasis-related apps. These differences were

found in the number of apps developed by a health-related technology company (50% in psoriasis apps and 6.7% in ankylosing spondylitis), nonhealth technology company (20% in psoriasis apps and 40% in ankylosing spondylitis), and in those apps where this information was not available (30% in psoriasis apps and 53% in ankylosing spondylitis; $P < .001$). No statistically significant differences were found between the rest of the recorded app characteristics and the type of IMIDs.

Table 3. Type of immune-mediated inflammatory diseases targeted by the apps.

Characteristics	Apps (N=243), n (%)
Ankylosing spondylitis	56 (23.0)
Crohn disease	93 (38.3)
Psoriatic arthritis	39 (16.0)
Psoriasis	50 (20.6)
Rheumatoid arthritis	111 (45.7)
Ulcerative colitis	89 (36.6)

App Functionalities

Analysis of the functionalities of the 243 apps revealed that 60.5% (n=147) had more than one functionality. A total of 135 apps offered information about lifestyle, nutrition, and physical exercise, and 102 apps offered information about the disease or its treatment (Table 4). We found that only 6 apps offered information specifically about biological therapies. Table 5

shows the main app functionalities related to the different IMIDs.

The mean number of functionalities per app was 2.1 (range 1-7). No statistically significant differences were found between the number of functionalities and the remaining variables, except for author affiliation. Apps developed by health-related technology companies presented a higher average number of functionalities.

Table 4. Functionalities of the apps.

Functionality	App (N=243), n (%)
General information about lifestyle, nutrition, exercise	135 (55.6)
Specific information about the disease or its treatment	102 (42.0)
Recording of symptoms or adverse events	51 (21.0)
Agenda/calendar	44 (18.1)
Medication reminder ^a	41 (16.9)
Recording of patient-reported outcomes	41 (16.9)
Monitoring the activity/severity of the disease	34 (14.7)
Contact with health care professionals	24 (9.9)
Social network	16 (6.6)
Other	36 (14.8)

^a27 of 41 made it possible to monitor adherence.

Table 5. Main app functionalities concerning the different immune-mediated inflammatory diseases.

	Information, n (%)	Recording of symptoms or adverse events, n (%)	Agenda/calendar, n (%)	Reminder medication, n (%)	Recording of PROs ^a , n (%)	Monitoring the activity/severity of the disease, n (%)	Contact with health professionals, n (%)	Social network, n (%)
AS ^b (n=56)	46 (82.1)	8 (14.3)	8 (14.3)	7 (12.5)	7 (12.5)	4 (7.1)	3 (5.4)	1 (1.8)
CD ^c (n=93)	68 (73.1)	21 (22.6)	24 (25.8)	19 (20.4)	19 (20.4)	15 (16.1)	9 (9.7)	5 (5.4)
PA ^d (n=39)	29 (74.5)	8 (20.5)	6 (15.4)	9 (23.1)	7 (17.9)	5 (12.8)	2 (5.1)	1 (2.6)
PS ^e (n=50)	40 (80.0)	14 (28.0)	8 (16.0)	8 (16.0)	11 (22.0)	4 (8.0)	9 (18.0)	3 (6.0)
RA ^f (n=111)	80 (72.1)	23 (20.7)	17 (15.3)	20 (18.0)	20 (18.0)	11 (9.9)	7 (6.3)	5 (4.5)
UC ^g (n=89)	64 (71.9)	22 (24.7)	24 (27.0)	20 (22.5)	16 (18.0)	14 (15.7)	10 (11.2)	7 (7.9)

^aPRO: patient-reported outcome.

^bAS: ankylosing spondylitis.

^cCD: Chron disease.

^dPA: psoriatic arthritis.

^ePS: psoriasis

^fRA: rheumatoid arthritis.

^gUC: ulcerative colitis.

Discussion

Principal Findings

We provide a comprehensive and unique review of IMID-related apps available for patients in the Play Store and Apple Store. The marketplace offers hundreds of solutions, thus making it difficult for the user to filter [17]. Health care professionals are unaware of the real situation of these apps and how they can help patients manage their disease. To the best of our knowledge, this is the only review published so far on the features and functionalities of apps for all IMIDs. The main findings show that functionalities and reliability were heterogeneous. Although most apps offer information, information on biological therapies is scarce. Few apps allow patient interaction, such as recording adverse events, setting alarms for medication, or recording PROs.

We analyzed 243 apps for the most important IMIDs and identified features that may contribute to our understanding of the current landscape and the development of future solutions. Most of the apps targeted rheumatoid arthritis and inflammatory bowel diseases (IBDs). This distribution is proportional to the prevalence of the IMIDs in real life [1]. Among the characteristics of the apps, we highlight the involvement of health care professionals in the development of 41.2% (n=100) of apps. However, 54.3% (n=132) of apps had not been updated in the last year, and the most frequent functionality was that of providing information.

The use of IMID-related apps is still in its infancy [11,18]. The medical fields with the most experience in apps are diabetes, mental health, and cardiovascular disease [7,17,19]. McKay et al [20] carried out a systematic review of apps related to one or more health conditions and found no IMID-related apps. In 2018, there were only 56 IBD-related apps available [21]. In a

cohort of 193 German patients with rheumatologic diseases, 91.2% regularly used a smartphone, and 68.4% believed that using medical apps could be beneficial for their health. However, only 4.1% of these patients used health-related apps, of which none were rheumatology-specific [8]. In a similar study performed on 575 French patients diagnosed with rheumatoid arthritis, only 4.7% used a health app [22].

In addition to the limited number of IMID apps, some of the negative characteristics such as a lack of robust scientific evidence prevent clinicians from using health-related apps or recommending them to patients [11,13,23]. It is estimated that only 38% of international consensus statements are completely covered by IBD-related apps focusing on patient education [4]. In a systematic review of self-management apps for persons with arthritic pain, Bhattarai et al [16] observed that a small number of arthritic pain apps offer a comprehensive self-management approach incorporating evidence-based strategies through the Stanford Arthritis Self-Management Program [16].

Currently, there is no specific legislation on the use of the apps and standards or official guidelines on app development. Similarly, there are no official repositories, mainly due to the fast-moving pace of mHealth [5,24]. The Food and Drug Administration launched the Digital Health Precertification Program (Pre-Cert Program) to evaluate mobile-based interventions in real-world conditions [7,25]. Despite the availability of validated tools for assessing the quality of apps, such as the Mobile App Rating Scale [26], there is minimal agreement on methods and the most appropriate criteria for this task [19]. Besides, few IMID-related apps have been developed or validated using these methodologies [14,27,28]. Author affiliation is an important characteristic when searching for a trustworthy app and has been a quality criterion in numerous studies [19]. However, most of the apps did not describe the credentials of the author or professionals who participated in the development. The fact that neither store requires qualifications or affiliations to be listed is one of the main reliability problems of health apps [10,17,29,30]. Thus, external sources, such as the developer's website if available, are needed. Even so, in 27.2% (n=66) of the apps we reviewed, we could not find any information about the author. We observed that more than half of the author affiliations were related to technology companies, although only 41% (n=100) were identified as involving some type of health professional. Only 7 of the 20 rheumatoid arthritis apps assessed by Luo et al [10] involved a health professional in content development. In another review of apps for medication management, 72.9% of apps were developed by technology companies, 2.1% by academic institutions, and 5.2% by other bodies (governments and nonprofit organizations), and there was insufficient information available about author affiliation in 17.1% [17]. Only 14.6% of apps were developed with the involvement of a health care professional. In a review of apps for rheumatic diseases, authors found that health care professionals were involved in the development of 40% of the apps [12]. Con and Cruz [4] reviewed 26 IBD-related apps and found that health professionals were involved in the development of 19.2% [4].

The level of the update is also a potential cause for concern. Given the rapid advance in IMID care, information must be constantly updated. In our study, we found that more than half of the apps had not been updated in the previous year, while in 25.1% (n=61), the last update had been over 2 years previously. In a review of 20 rheumatoid arthritis apps, only 12 had been updated in the previous year [10]. Concerning other diseases, Collado-Borrell et al [31] observed that 52.4% of cancer-related apps had updated their software within the previous year. We observed that the apps available in the Play Store were more recently updated than those in the App Store. These results contrast with those of studies that have shown the average quality of health apps to be better on iOS than on Android [32].

On the other hand, one of the advantages of these tools is that most health-related apps are free [4,17,31,33], thus favoring their accessibility. Consistent with other studies, we observed that 82.3% (n=200) of IMID-related apps were free [10,17,31,33]. However, in IMID apps requiring payment, the cost is slightly higher (US \$10.30) than the average observed in other reviews (US \$1.01-4.74) [4,9,10,17,31,33]. Concerning the country of development, the most frequent was the United States, followed by India. Other app review studies have shown a similar distribution [33], likely due to the potential of information and communication technology companies in these countries.

Many of the health-related apps emerged as tools for consulting information, and this is still the most frequent functionality, although 60.5% (n=147) of the apps analyzed had more than one functionality. Our results are consistent with those of other studies and highlight that most apps offer passive functionalities such as information about the disease, treatments, lifestyle, nutrition, and physical exercise. Medication information was the feature most desired by a cohort of 193 patients with rheumatic disease [8]. Mollard et al [11] observed that 25% of apps available for rheumatoid arthritis were exclusively for patient education. Con and Cruz [4] observed that 50% of IBD apps had exclusively diary functionalities that made it possible to track several activities or records, although there was considerable heterogeneity regarding the detail with which symptoms were able to be recorded. In that study, 34.6% of apps only provided information, 11.5% recorded PROs, and 7.7% had two reminder systems. Only 1 app allowed communication between patients and professionals. Similar results were obtained in a review of rheumatoid arthritis apps, in which 50% only allowed symptom tracking and 25% had only provided information [10]. Tabi et al [17] analyzed 328 apps for medication management and observed that 41.2% could share data and reports, although no app allowed bidirectional communication with health professionals.

Several studies have shown the usefulness of apps for improving adherence [17,34]. This feature is even more important in drugs that are not administered daily, such as biological therapies, as delays in administration may be more frequent. For instance, 20% to 40% of patients with IBD are nonadherent. Alerts or reminders to take medication are the most frequent feature of apps and improve adherence. In general, apps are the preferred method for reminding patients to take their medication regularly [8]. However, in our review, only 16.9% (n=41) of apps have

this functionality, and only 11.1% (n=27) enable medication adherence to be monitored.

Encouraging enthusiastic patients to engage with apps regularly is another barrier [11]. For example, many patients download an app for the sole purpose of obtaining health information. However, after initial use of the app, there is unlikely to be continual engagement once the desired information is obtained [11]. Long-term engagement can be problematic, with app interventions progressively becoming less effective over time [7]. Some functionalities involving patient participation, such as sending messages and gamification, can also improve long-term engagement [19]. Gamification is a specific characteristic of mHealth interventions that could favor behavioral change and increase user engagement. However, we only found 1 app with the functionality of gamification in our review.

Limitations and Strengths

A possible limitation is that we did not assess the privacy and confidentiality of apps since this information is not available in most cases. The relationship between the use of health apps and privacy risk is understudied [10]. Most reviews and studies that assess app quality do not analyze this aspect [19,20]. In fact, the number of health apps that define privacy policies is

low [35]. Nevertheless, to our knowledge, this is the first review to cover most IMIDs and analyze their characteristics in detail. We believe that our review can provide health professionals with a real-world picture of the apps available for patients with IMIDs and help them to identify strengths and weaknesses. We have observed that there are a large number of free apps in English available for patients diagnosed with IMIDs, especially for consulting information. However, when recommending apps or instructing patients to search for them, health care professionals should be aware of some critical factors. Professionals and patients should check the author's affiliation and the update date, as there are many IMID apps with nonproven developers and without updates. Future legislation for apps should focus specifically on bridging these gaps.

Conclusions

This review provides health professionals with an overview of available IMID-related apps and their characteristics. IMID-related apps continue to be heterogeneous in terms of their functionalities and reliability. Apps may act as a useful complement to IMID care, especially in patient education, as this is their most frequent functionality. However, more than half of the IMID apps had not been developed by health care professionals or updated in the previous year.

Conflicts of Interest

None declared.

References

1. El-Gabalawy H, Guenther LC, Bernstein CN. Epidemiology of immune-mediated inflammatory diseases: incidence, prevalence, natural history, and comorbidities. *J Rheumatol Suppl* 2020 May;85:2-10. [doi: [10.3899/jrheum.091461](https://doi.org/10.3899/jrheum.091461)] [Medline: [20436161](https://pubmed.ncbi.nlm.nih.gov/20436161/)]
2. Baker KF, Isaacs JD. Novel therapies for immune-mediated inflammatory diseases: what can we learn from their use in rheumatoid arthritis, spondyloarthritis, systemic lupus erythematosus, psoriasis, Crohn's disease and ulcerative colitis? *Ann Rheum Dis* 2018 Feb;77(2):175-187. [doi: [10.1136/annrheumdis-2017-211555](https://doi.org/10.1136/annrheumdis-2017-211555)] [Medline: [28765121](https://pubmed.ncbi.nlm.nih.gov/28765121/)]
3. 53% of EU citizens sought health information online. European Commission. 2020. URL: <https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20200327-1> [accessed 2021-04-15]
4. Con D, De Cruz P. Mobile phone apps for inflammatory bowel disease self-management: a systematic assessment of content and tools. *JMIR Mhealth Uhealth* 2016 Feb 01;4(1):e13 [FREE Full text] [doi: [10.2196/mhealth.4874](https://doi.org/10.2196/mhealth.4874)] [Medline: [26831935](https://pubmed.ncbi.nlm.nih.gov/26831935/)]
5. Sharma A, Harrington RA, McClellan MB, Turakhia MP, Eapen ZJ, Steinhubl S, et al. Using digital health technology to better generate evidence and deliver evidence-based care. *J Am Coll Cardiol* 2018 Jun 12;71(23):2680-2690 [FREE Full text] [doi: [10.1016/j.jacc.2018.03.523](https://doi.org/10.1016/j.jacc.2018.03.523)] [Medline: [29880129](https://pubmed.ncbi.nlm.nih.gov/29880129/)]
6. Morillo-Verdugo R, Margusino-Framiñán L, Monte-Boquet E, Morell-Baladrón A, Barreda-Hernández D, Rey-Piñeiro XM, et al. Spanish Society of Hospital Pharmacy Position Statement on Telepharmacy: recommendations for its implementation and development. *Farm Hosp* 2020 Jul 01;44(4):174-181. [doi: [10.7399/fh.11515](https://doi.org/10.7399/fh.11515)] [Medline: [32646349](https://pubmed.ncbi.nlm.nih.gov/32646349/)]
7. Jiménez-Muñoz L, Gutiérrez-Rojas L, Porras-Segovia A, Courtet P, Baca-García E. Mobile applications for the management of chronic physical conditions: a systematic review. *Intern Med J* 2022 Jan;52(1):21-29. [doi: [10.1111/imj.15081](https://doi.org/10.1111/imj.15081)] [Medline: [33012045](https://pubmed.ncbi.nlm.nih.gov/33012045/)]
8. 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](https://doi.org/10.2196/19661)] [Medline: [32678796](https://pubmed.ncbi.nlm.nih.gov/32678796/)]
9. Brewer AC, Endly DC, Henley J, Amir M, Sampson BP, Moreau JF, et al. Mobile applications in dermatology. *JAMA Dermatol* 2013 Nov;149(11):1300-1304. [doi: [10.1001/jamadermatol.2013.5517](https://doi.org/10.1001/jamadermatol.2013.5517)] [Medline: [24067948](https://pubmed.ncbi.nlm.nih.gov/24067948/)]
10. Luo D, Wang P, Lu F, Elias J, Sparks JA, Lee YC. Mobile apps for individuals with rheumatoid arthritis: a systematic review. *J Clin Rheumatol* 2019 Apr;25(3):133-141. [doi: [10.1097/RHU.0000000000000800](https://doi.org/10.1097/RHU.0000000000000800)] [Medline: [29933327](https://pubmed.ncbi.nlm.nih.gov/29933327/)]
11. Mollard E, Michaud K. Mobile apps for rheumatoid arthritis: opportunities and challenges. *Rheum Dis Clin North Am* 2019 May;45(2):197-209. [doi: [10.1016/j.rdc.2019.01.011](https://doi.org/10.1016/j.rdc.2019.01.011)] [Medline: [30952393](https://pubmed.ncbi.nlm.nih.gov/30952393/)]

12. Najm A, Gossec L, Weill C, Benoist D, Berenbaum F, Nikiphorou E. Mobile health apps for self-management of rheumatic and musculoskeletal diseases: systematic literature review. *JMIR Mhealth Uhealth* 2019 Nov 26;7(11):e14730 [FREE Full text] [doi: [10.2196/14730](https://doi.org/10.2196/14730)] [Medline: [31769758](https://pubmed.ncbi.nlm.nih.gov/31769758/)]
13. Yin AL, Hachuel D, Pollak JP, Scherl EJ, Estrin D. Digital health apps in the clinical care of inflammatory bowel disease: scoping review. *J Med Internet Res* 2019 Aug 19;21(8):e14630 [FREE Full text] [doi: [10.2196/14630](https://doi.org/10.2196/14630)] [Medline: [31429410](https://pubmed.ncbi.nlm.nih.gov/31429410/)]
14. Kwan YH, Ong WJ, Xiong M, Leung YY, Phang JK, Wang CTM, et al. Evaluation of mobile apps targeted at patients with spondyloarthritis for disease monitoring: systematic app search. *JMIR Mhealth Uhealth* 2019 Oct 28;7(10):e14753 [FREE Full text] [doi: [10.2196/14753](https://doi.org/10.2196/14753)] [Medline: [31661080](https://pubmed.ncbi.nlm.nih.gov/31661080/)]
15. Luckett T, Bhattarai P, Phillips J, Agar M, Currow D, Krastev Y, et al. Advance care planning in 21st century Australia: a systematic review and appraisal of online advance care directive templates against national framework criteria. *Aust Health Rev* 2015 Nov;39(5):552-560. [doi: [10.1071/AH14187](https://doi.org/10.1071/AH14187)] [Medline: [26027723](https://pubmed.ncbi.nlm.nih.gov/26027723/)]
16. Bhattarai P, Newton-John TRO, Phillips JL. Quality and usability of arthritic pain self-management apps for older adults: a systematic review. *Pain Med* 2018 Mar 01;19(3):471-484. [doi: [10.1093/pm/pnx090](https://doi.org/10.1093/pm/pnx090)] [Medline: [28541464](https://pubmed.ncbi.nlm.nih.gov/28541464/)]
17. Tabi K, Randhawa AS, Choi F, Mithani Z, Albers F, Schnieder M, et al. Mobile apps for medication management: review and analysis. *JMIR Mhealth Uhealth* 2019 Sep 11;7(9):e13608 [FREE Full text] [doi: [10.2196/13608](https://doi.org/10.2196/13608)] [Medline: [31512580](https://pubmed.ncbi.nlm.nih.gov/31512580/)]
18. Bossuyt P, Pouillon L, Bonnaud G, Danese S, Peyrin-Biroulet L. E-health in inflammatory bowel diseases: more challenges than opportunities? *Dig Liver Dis* 2017 Dec;49(12):1320-1326. [doi: [10.1016/j.dld.2017.08.026](https://doi.org/10.1016/j.dld.2017.08.026)] [Medline: [28899622](https://pubmed.ncbi.nlm.nih.gov/28899622/)]
19. Agarwal P, Gordon D, Griffith J, Kithulegoda N, Witteman HO, Sacha Bhatia R, et al. Assessing the quality of mobile applications in chronic disease management: a scoping review. *NPJ Digit Med* 2021 Mar 10;4(1):46. [doi: [10.1038/s41746-021-00410-x](https://doi.org/10.1038/s41746-021-00410-x)] [Medline: [33692488](https://pubmed.ncbi.nlm.nih.gov/33692488/)]
20. McKay FH, Cheng C, Wright A, Shill J, Stephens H, Uccellini M. Evaluating mobile phone applications for health behaviour change: a systematic review. *J Telemed Telecare* 2018 Jan;24(1):22-30. [doi: [10.1177/1357633X16673538](https://doi.org/10.1177/1357633X16673538)] [Medline: [27760883](https://pubmed.ncbi.nlm.nih.gov/27760883/)]
21. George LA, Cross RK. Remote monitoring and telemedicine in IBD: are we there yet? *Curr Gastroenterol Rep* 2020 Feb 10;22(3):12. [doi: [10.1007/s11894-020-0751-0](https://doi.org/10.1007/s11894-020-0751-0)] [Medline: [32040650](https://pubmed.ncbi.nlm.nih.gov/32040650/)]
22. Magnol M, Eleonore B, Claire R, Castagne B, Pugibet M, Lukas C, et al. Use of eHealth by patients with rheumatoid arthritis: observational, cross-sectional, multicenter study. *J Med Internet Res* 2021 Jan 29;23(1):e19998 [FREE Full text] [doi: [10.2196/19998](https://doi.org/10.2196/19998)] [Medline: [33512320](https://pubmed.ncbi.nlm.nih.gov/33512320/)]
23. Grainger R, Townsley H, White B, Langlotz T, Taylor WJ. Apps for people with rheumatoid arthritis to monitor their disease activity: a review of apps for best practice and quality. *JMIR Mhealth Uhealth* 2017 Feb 21;5(2):e7 [FREE Full text] [doi: [10.2196/mhealth.6956](https://doi.org/10.2196/mhealth.6956)] [Medline: [28223263](https://pubmed.ncbi.nlm.nih.gov/28223263/)]
24. Chan CX, Markowitz O. APPLYing knowledge: evidence for and regulation of mobile apps for dermatologists. *Cutis* 2020 Aug;106(2):76-78. [doi: [10.12788/cutis.0036](https://doi.org/10.12788/cutis.0036)] [Medline: [32941561](https://pubmed.ncbi.nlm.nih.gov/32941561/)]
25. Lee TT, Kesselheim AS. U.S. Food and Drug Administration Precertification Pilot Program for Digital Health Software: weighing the benefits and risks. *Ann Intern Med* 2018 May 15;168(10):730-732. [doi: [10.7326/M17-2715](https://doi.org/10.7326/M17-2715)] [Medline: [29632953](https://pubmed.ncbi.nlm.nih.gov/29632953/)]
26. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. *JMIR Mhealth Uhealth* 2015 Mar 11;3(1):e27 [FREE Full text] [doi: [10.2196/mhealth.3422](https://doi.org/10.2196/mhealth.3422)] [Medline: [25760773](https://pubmed.ncbi.nlm.nih.gov/25760773/)]
27. Bearne LM, Sekhon M, Grainger R, La A, Shamali M, Amirova A, et al. Smartphone apps targeting physical activity in people with rheumatoid arthritis: systematic quality appraisal and content analysis. *JMIR Mhealth Uhealth* 2020 Jul 21;8(7):e18495 [FREE Full text] [doi: [10.2196/18495](https://doi.org/10.2196/18495)] [Medline: [32706727](https://pubmed.ncbi.nlm.nih.gov/32706727/)]
28. Dantas LO, Carvalho C, Prando BC, McAlindon TE, da Silva Serrão PRM. Mobile health technologies for the management of rheumatic diseases: a systematic review of online stores in Brazil. *Clin Rheumatol* 2021 Jul;40(7):2601-2609. [doi: [10.1007/s10067-020-05561-y](https://doi.org/10.1007/s10067-020-05561-y)] [Medline: [33389312](https://pubmed.ncbi.nlm.nih.gov/33389312/)]
29. Brouard B, Bardo P, Bonnet C, Mounier N, Vignot M, Vignot S. Mobile applications in oncology: is it possible for patients and healthcare professionals to easily identify relevant tools? *Ann Med* 2016 Nov;48(7):509-515. [doi: [10.1080/07853890.2016.1195010](https://doi.org/10.1080/07853890.2016.1195010)] [Medline: [27348761](https://pubmed.ncbi.nlm.nih.gov/27348761/)]
30. Subhi Y, Bube SH, Rolskov Bojsen S, Skou Thomsen AS, Konge L. Expert involvement and adherence to medical evidence in medical mobile phone apps: a systematic review. *JMIR Mhealth Uhealth* 2015 Jul 27;3(3):e79 [FREE Full text] [doi: [10.2196/mhealth.4169](https://doi.org/10.2196/mhealth.4169)] [Medline: [26215371](https://pubmed.ncbi.nlm.nih.gov/26215371/)]
31. Collado-Borrell R, Escudero-Vilaplana V, Ribed-Sánchez A, Ibáñez-García S, Herranz-Alonso A, Sanjurjo-Sáez M. Smartphone applications for cancer patients; what we know about them? *Fam Hosp* 2016 Jan 01;40(1):25-35 [FREE Full text] [doi: [10.7399/fh.2016.40.1.8993](https://doi.org/10.7399/fh.2016.40.1.8993)] [Medline: [26882831](https://pubmed.ncbi.nlm.nih.gov/26882831/)]
32. Amor-García M, Collado-Borrell R, Escudero-Vilaplana V, Melgarejo-Ortuño A, Herranz-Alonso A, Arranz Arija J, et al. Assessing apps for patients with genitourinary tumors using the Mobile Application Rating Scale (MARS): systematic search in app stores and content analysis. *JMIR Mhealth Uhealth* 2020 Jul 23;8(7):e17609 [FREE Full text] [doi: [10.2196/17609](https://doi.org/10.2196/17609)] [Medline: [32706737](https://pubmed.ncbi.nlm.nih.gov/32706737/)]

33. 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](https://doi.org/10.2196/20334)] [Medline: [32614777](https://pubmed.ncbi.nlm.nih.gov/32614777/)]
34. Haase J, Farris KB, Dorsch MP. Mobile applications to improve medication adherence. *Telemed J E Health* 2017 Feb;23(2):75-79. [doi: [10.1089/tmj.2015.0227](https://doi.org/10.1089/tmj.2015.0227)] [Medline: [27248315](https://pubmed.ncbi.nlm.nih.gov/27248315/)]
35. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. *J Am Med Inform Assoc* 2015 Apr;22(e1):e28-e33. [doi: [10.1136/amiajnl-2013-002605](https://doi.org/10.1136/amiajnl-2013-002605)] [Medline: [25147247](https://pubmed.ncbi.nlm.nih.gov/25147247/)]

Abbreviations

IBD: inflammatory bowel disease

IMID: immune-mediated inflammatory disease

mHealth: mobile health

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

PRO: patient-reported outcome

Edited by A Mavragani; submitted 09.06.21; peer-reviewed by R Collado-Borrell, M Salimi, C Cheng; comments to author 30.07.21; revised version received 03.08.21; accepted 23.12.21; published 04.03.22.

Please cite as:

Romero-Jimenez R, Escudero-Vilaplana V, Chamorro-De-Vega E, Ais-Larisgoitia A, Lobato Matilla ME, Herranz-Alonso A, Sanjurjo M

The Characteristics and Functionalities of Mobile Apps Aimed at Patients Diagnosed With Immune-Mediated Inflammatory Diseases: Systematic App Search

J Med Internet Res 2022;24(3):e31016

URL: <https://www.jmir.org/2022/3/e31016>

doi:[10.2196/31016](https://doi.org/10.2196/31016)

PMID:[35254286](https://pubmed.ncbi.nlm.nih.gov/35254286/)

©Rosa Romero-Jimenez, Vicente Escudero-Vilaplana, Esther Chamorro-De-Vega, Arantza Ais-Larisgoitia, Maria Elena Lobato Matilla, Ana Herranz-Alonso, Maria Sanjurjo. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 04.03.2022. 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.

Original Paper

An mHealth Intervention to Reduce the Packing of Discretionary Foods in Children's Lunch Boxes in Early Childhood Education and Care Services: Cluster Randomized Controlled Trial

Nicole Pearson^{1,2,3,4}, BSc; Meghan Finch^{1,2,3,4}, PhD; Rachel Sutherland^{1,2,3,4}, PhD; Melanie Kingsland^{1,2,3,4}, PhD; Luke Wolfenden^{1,2,3,4}, PhD; Taya Wedesweiler¹, BHealthPhysEd; Vanessa Herrmann¹, DipNutr; Sze Lin Yoong^{1,2,3,4}, PhD

¹Hunter New England Population Health, Wallsend, Australia

²School of Medicine and Public Health, University of Newcastle, Callaghan, Australia

³Hunter Medical Research Institute, Newcastle, Australia

⁴Priority Research Centre for Health Behaviour, University of Newcastle, Callaghan, Australia

Corresponding Author:

Nicole Pearson, BSc

Hunter New England Population Health

Longworth Avenue

Wallsend, 2308

Australia

Phone: 61 02 49246031 ext 46031

Fax: 61 02 4924 6490

Email: Nicole.Pearson@health.nsw.gov.au

Abstract

Background: Interventions in early childhood education and care (ECEC) services have the potential to improve children's diet at the population level.

Objective: This study aims to test the efficacy of a mobile health intervention in ECEC services to reduce parent packing of foods high in saturated fat, sugar, and sodium (discretionary foods) in children's (aged 3-6 years) lunch boxes.

Methods: A cluster randomized controlled trial was undertaken with 355 parent and child dyads recruited by phone and in person from 17 ECEC services (8 [47%] intervention and 9 [53%] control services). Parents in the intervention group received a 10-week fully automated program targeting barriers to packing healthy lunch boxes delivered via an existing service communication app. The program included weekly push notifications, within-app messages, and links to further resources, including websites and videos. The control group did not receive any intervention. The primary outcomes were kilojoules from discretionary foods and associated nutrients (saturated fat, free sugars, and sodium) packed in children's lunch boxes. Secondary outcomes included consumption of kilojoules from discretionary foods and related nutrients and the packing and consumption of serves of discretionary foods and core food groups. Photography and weights of foods in children's lunch boxes were recorded by trained researchers before and after the trial to assess primary and secondary outcomes. Outcome assessors were blinded to service allocation. Feasibility, appropriateness, and acceptability were assessed via an ECEC service manager survey and a parent web-based survey. Use of the app was assessed via app analytics.

Results: Data on packed lunch box contents were collected for 88.8% (355/400) of consenting children at baseline and 84.3% (337/400) of children after the intervention. There was no significant difference between groups in kilojoule from discretionary foods packed (77.84 kJ, 95% CI -163.49 to 319.18; $P=.53$) or the other primary or secondary outcomes. The per-protocol analysis, including only data from children of parents who downloaded the app, also did not find any statistically significant change in primary (-1.98 kJ, 95% CI -343.87 to 339.90; $P=.86$) or secondary outcomes. Approximately 61.8% (102/165) of parents in the intervention group downloaded the app, and the mean service viewing rate of weekly within-app messages was 26% (SD 14.9). Parents who responded to the survey and participating services agreed that it was appropriate to receive lunch box information via the app (40/50, 80% and 6/8, 75%, respectively).

Conclusions: The intervention was unable to demonstrate an impact on kilojoules or associated nutrients from discretionary foods packed in children's lunch boxes. Low app downloads and program message views indicate a need to explore how to improve factors related to implementation before further testing similar mobile health interventions in this setting.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12618000133235; <https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=374379>

(*J Med Internet Res* 2022;24(3):e27760) doi:[10.2196/27760](https://doi.org/10.2196/27760)

KEYWORDS

nutrition; mHealth; child; preschool; parents

Introduction

Background

Poor diet characterized by excessive intake of foods high in sugar, salt, and sodium and low intake of core foods such as fruits and vegetables is associated with a higher risk of chronic disease [1]. The consumption of foods high in saturated fat, sugar, or salt exceeds optimal levels worldwide [2], and in countries such as the United States, the United Kingdom, and Australia, excessive consumption of such nutrients has been reported across all age groups [3-5]. In Australia, foods high in saturated fat, sugar, or sodium, and low in nutrients such as fiber (from here on referred to as *discretionary foods*) [6], contribute up to 38% of the total daily energy intake of a child aged 4 to 8 years [5]. Given this, international and national dietary guidelines recommend limiting the consumption of discretionary foods [7-10].

Early childhood education and care (ECEC) services constitute an ideal setting for reaching and engaging families to influence lifelong eating habits [11,12]. In the United States, the United Kingdom, and Australia, up to 80% of children attend some kind of formal ECEC service in the year before school [13-15], with an average attendance of 2 to 3 days per week [15]. A significant number of ECEC services in these countries require children to bring food from home in a lunch box (approximately 30%-50%) [13,14,16]. Australian studies report that between 50% and 60% of lunch boxes include discretionary foods [17,18], with one reporting an average of 2 serves included per lunch box [18]. As such, there is potential to improve children's diet by targeting the packing of discretionary foods in lunch boxes.

A systematic review of interventions to improve the contents of lunch boxes identified just 3 randomized controlled trials that examined interventions to improve the packing of discretionary foods within the ECEC setting [17,19,20]. Of the 3 trials, 2 reported significant results for reducing the packing of discretionary foods [17,19]. Both were intensive multicomponent interventions, including teacher education, child curriculum, and face-to-face education sessions with parents [17,19]. Despite the success of these interventions, evidence suggests that the sustainability and scalability of such intensive designs may be limited [21,22]. In particular, face-to-face parent-related components have been identified as difficult for parents to attend because of time and travel burden [23] and challenging for ECEC services to implement [21].

The use of digital health-delivered interventions has been proposed as a way of overcoming barriers to engaging parents in nutrition interventions, as well as enhancing scalability [24-26]. Digital health interventions (DHIs), in particular the

use of mobile phone apps or mobile health (mHealth) interventions, have been identified as an effective way of providing health-related information to parents [27] and influencing child health outcomes [25,28-30]. In addition, mobile apps designed to facilitate parent ECEC service communications are increasingly available for use by ECEC services in Australia, offering an alternative to face-to-face delivery of nutrition interventions by drawing on existing digital platforms [26]. Using existing apps to deliver nutrition interventions has also been recommended to capitalize on technologies developed by experts in user-centered design and with established commercial appeal [26]. Therefore, integrating a lunch box intervention into an existing ECEC parent communication app shows promise as a novel way of delivering digital content and engaging with parents.

Objective

To the authors' knowledge, no trials have been conducted to assess whether mHealth interventions can support the packing of healthy lunch boxes for children attending ECEC services. To address this gap in research evidence, the aim of this study is to determine the efficacy of using an mHealth intervention versus no intervention in ECEC services to reduce the packing of discretionary foods in children's lunch boxes. We hypothesize that as a result of the intervention, lunch boxes of children with parents in the intervention group would achieve a mean reduction of 123 kJ from packed discretionary foods relative to the control. Additional outcomes include the intervention's impact on children's consumption of food, parental use of the app, intervention fidelity, and other process evaluation measures.

Methods

Ethics Approval

Ethical approval was obtained from the Hunter New England (HNE) human research ethics committee (06/07/26/4.04) and ratified by the University of Newcastle ethics committee (H-2008-0343). The trial is reported in line with the CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online Telehealth) reporting guidelines ([Multimedia Appendix 1](#)) [31], and a protocol has been previously published [32].

Study Design and Setting

The study used a cluster randomized controlled trial design and targeted the parents of children attending center-based ECEC services (ie, long day care and preschools). The ECEC services were located in the HNE Local Health District of New South Wales (NSW), Australia. ECEC services were randomized to a 10-week intervention group or no intervention control group. Within NSW, long daycare services can provide center-based

education and care for children aged from 6 weeks to 6 years for ≥ 8 hours per day. Preschools typically enroll children aged between 3 and 6 years and provide care for 6 to 8 hours per day [33]. In 2016, approximately 920,370 people were reported to live in the HNE area, with 51,900 of these being children aged 0 to 4 years [34]. The area includes major metropolitan centers and inner regional communities, with a small proportion (14%) of people in remote communities [35].

Sample

The sampling frame comprised ECEC services from the region that required parents to provide food for consumption at the service (ie, lunch box services). The list of ECEC services that required parents to provide food was obtained from a government database and represented approximately 53% of all ECEC services in the region [16]. Of these services, existing users of the app required for the intervention and those not using any app (ie, services able to commence using the intervention app) were identified using previous, unpublished data on app use by services in the region.

Recruitment and Eligibility

Recruitment for the trial occurred in 2 phases. Initially, ECEC services were eligible to participate in the trial if they enrolled children aged 3 to 6 years and were existing users of the designated parent communication app. As this eligibility criterion did not result in adequate ECEC services being recruited, phase 2 recruitment extended the eligibility criteria to include ECEC services not yet using the app but willing to commence using the app for the trial. For both phases, recruitment involved research assistants posting and emailing information statements and consent forms to ECEC services outlining the study, data collection procedures, and requesting participation. Written consent was obtained from service managers. Parents or carers (hereafter referred to as *parents*) of children aged 3 to 6 years were eligible to participate if their child attended during the designated days of data collection and if they used or indicated a willingness to download the intervention app on the consent form. To obtain parental consent for participation in the study, ECEC service staff distributed hard copies of parent information statements and consent forms. Parents could also consent on the day of the data collection.

Random Allocation and Blinding

Before baseline data collection, ECEC services were randomly allocated to the intervention group or no intervention control group in a 1:1 ratio by a statistician independent of the trial using a computerized random number generator. Before randomization, ECEC services were stratified by rural location and socioeconomic status (SES) of the service, as evidence indicates that these factors are associated with family dietary intake [36,37]. To ensure equity of access to the intervention, ECEC services were also stratified by those with high numbers of Aboriginal child enrollments (defined as those with $>10\%$ Aboriginal children enrolled). Stratification by this number of factors was deemed appropriate for the sample size [38]. Owing to the nature of the intervention, ECEC services and parents were not blinded to the intervention; however, outcome assessors were blinded to the service allocation.

Intervention Development

Details of the intervention and its development, including the application of theory, cultural, and other stakeholder consultation processes, have been published elsewhere [32]. Briefly, the intervention was based on an mHealth lunch box intervention originally designed for primary (elementary) schools by the research team, which reported promising pilot data on the packing of discretionary items in school [39,40]. Similar to the original intervention, the behavior change wheel (BCW) and COM-B (Capability, Opportunity, Motivation, and Behavior) model were used to inform the current intervention [41]. The BCW framework is based on 19 theories of health behavior change and facilitates the systematic development of behavior change interventions. The COM-B model is the behavior system behind the BCW framework, which supports the identification of essential conditions for changes, including capability, opportunity, and motivation [41]. The resulting mHealth intervention, *SWAP IT for Childcare*, comprised 7 behavior change techniques targeting 8 barriers to packing healthy lunch boxes. (Multimedia Appendix 2). Barriers were identified from the literature and formative interviews with a convenience sample of parents from 3 local ECEC services. The adaptation process was then guided by the Framework for Reporting Adaptations and Modifications–Enhanced (Multimedia Appendix 3) [42].

Intervention App

The app used for the intervention (Skoolbag) was an existing app designed to be available for use by schools or ECEC services to communicate with parents. Commonly, ECEC services use the app to share parent newsletters, reminders, and other service information. The app had not been previously used to deliver nutrition interventions in the ECEC setting.

Intervention Strategies

The *SWAP IT for Childcare* intervention comprised 3 components.

Provision of Weekly Push Notifications and Within-App Messages

A total of 11 push notifications messages were delivered over 10 weeks (1 per week, with an additional introductory message delivered in week 1). The push notifications were sent by the app provider. Each push notification alerted users to a within-app message, which aimed to target parent barriers to packing healthy lunch boxes. The message was accessible via the push notification or could be accessed as part of the static content within the app. A summary of each push notification, within-app message, and the behavior change techniques is available in the published protocol [32].

Provision of SWAP IT Options and Supporting Resources Via the App

Several of the within-app messages provided a weblink to *SWAP IT Options*—a comprehensive list of foods suitable for packing in the lunch box. The web-based list was developed by dietitians on the research team with expertise working within the setting. The lists were divided into sections (sweet snacks, savory snacks, and lunch foods) with drop-down tabs to explore foods

recommended as *swap from* and *swap to*. Links to other supporting information relevant to each message were also provided, including fact sheets, short videos, and website links.

ECEC Service Endorsement of Program

To encourage initial and ongoing engagement with the app content, intervention ECEC service managers were asked to endorse the *SWAP IT* program and demonstrate support to parents. To do this, the research team provided 2 templates to service managers to communicate with parents during the course of the trial, which service managers could deliver via the app or by email. ECEC services were asked to deliver the communications the week before the intervention and midintervention (weeks 5-6).

Service Implementation Strategies to Increase App Uptake and Engagement

Although not an implementation trial, to maximize uptake and engagement with the intervention, some implementation strategies were undertaken [43].

Use Financial Strategies and Train Stakeholders

Before the intervention, ECEC services that had not used the app before the trial (intervention: 6/17, 35%; control: 5/17, 29%) had their access paid for by the research team during the trial period. Remote training was provided to new ECEC services on how to use the app by the app provider.

Engage Consumers

On recruitment, parents were supported to gain access to the app if not already doing so. Multiple strategies were used to maximize the number of parents on the app before and during the intervention. This included the research team providing emails to service managers to send to parents, as well as researchers sending emails sent directly to consenting parents asking them to download the app. Emails included step-by-step instructions on how to download the app and were distributed via printed flyers to ECEC services.

Control

The ECEC services allocated to the control group received no intervention. ECEC services had access to the app to use for their own communications but did not receive any of the intervention strategies or content.

Sample Size and Power Calculations

This study aimed to recruit 390 children from 18 ECEC services. Allowing a 15% attrition rate, this would enable the detection of a mean difference of 123 (SD 200) kJ in the primary outcome (kilojoules from discretionary foods), with an α of .01 (adjusting for multiple outcomes using Bonferroni adjustments) and an estimated intraclass correlation coefficient of 0.1 with 80% power. Such an energy reduction would represent approximately a quarter of a serve reduction in packed discretionary foods and could be expected to result in the detection of approximately 2.2 g less sugar, 0.6 g less saturated fat, and 44 mg less sodium [32].

Measures and Data Collection Procedures

Overview

Service baseline data were collected between May 2018 and July 2018. The 10-week intervention was delivered between July and September 2018. Postintervention data were collected for child lunch box outcomes during October and November 2018, and parent data were collected from October to December 2018.

The primary outcome measures were mean energy (kJ) provided by discretionary foods and mean energy (kJ), saturated fat (g), free sugars (mg), and sodium (mg) provided by all foods packed in the lunch box. Secondary outcomes were mean energy from discretionary foods (kJ) and mean energy (kJ), saturated fat (g), free sugars (mg), and sodium (mg) from all foods consumed by children from their lunch box. Other secondary outcomes were the number of serves of core food groups (bread and cereals, fruits, vegetables, dairy, meat, and meat alternatives) and number of discretionary food serves packed and consumed by children.

Packing and Consumption of Foods from Lunch Boxes

To assess the effectiveness of the intervention, data on food contents packed and consumed from lunch boxes were collected by trained research assistants, on one day, via weighed food records before the child's first meal and at the end of the child's last meal. ECEC services were asked not to inform parents of the day of data collection to reduce performance bias on lunch box packing behaviors. Weighed food records were used as evidence indicates this is the most accurate way of capturing portion sizes and quantities eaten [44,45]. On the day of data collection, the children were instructed to leave all uneaten food in the lunch box. Lunch box contents were also recorded on written standardized forms and digitally photographed to verify weighed records and identify inedible waste. Food record data were extracted by trained dietitians and entered into a nutrition analysis database (FoodWorks) [46]. Consumption data were derived by subtracting the weight of the food left at the completion of the last meal from the weight of food packed before meals. FoodWorks was used to produce nutrient data and the core food group serves data for both packed and consumed amounts of food. The software uses core food group classifications and serve sizes consistent with the Australian Guide to Healthy Eating [6]. For foods that were homemade, an appropriate standard recipe was sourced from within the FoodWorks database. As FoodWorks does not have the function to calculate discretionary food serves, this was undertaken manually by trained dietitians, using definitions based on the Australian Guide to Healthy Eating [6]. If further guidance was required, discretionary food classifications were sourced from the National Nutrition Survey food classification databases [47] or consensus from the research team. Further details of the FoodWorks data extraction processes and analysis are reported elsewhere [48].

Parent and Service Characteristics

Parent-child dyad characteristics were collected via consent forms and as part of a wider web-based survey sent via parent email before the intervention. Service operational characteristics

were collected via service manager pen and paper surveys completed at baseline. During a site visit, the service nutrition context was measured by trained observers using a modified version of the Environment and Policy Assessment Observation (EPAO) tool. The tool was adapted to include only nutrition items from the original EPAO tool relevant to ECEC services where foods are brought from home (ie, items related to services that provide food were omitted). Additional items were added related to staff monitoring of lunch boxes for discretionary foods, staff actions taken as a result of discretionary foods in lunch boxes, and the presence of specific lunch box guidelines at the service [49].

Process Evaluation

Intervention fidelity was captured via app analytics and through a service manager, self-completed written record. App analytics were used to access data on app downloads and the *SWAP IT* program use. Parent and service measures of acceptability and feasibility of the *SWAP IT* program were undertaken using a web-based survey and service manager pen and paper survey. Acceptability and feasibility questions were modified versions of items taken from the Acceptability of Intervention Measure and the Feasibility of Intervention Measure [50]. Questions to determine if any cointerventions occurred during the trial and capture any adverse events as a result of the intervention were included in the service manager pen and paper survey. Further details of the process data collection methods and items used within the surveys are provided in [Multimedia Appendix 4](#) [50].

Statistical Analysis

A statistician independent of the study performed all analyses using SAS (version 9.4) statistical software. Descriptive statistics were generated for service, parent, and child characteristics. The Australian Statistical Geography Standard Remoteness Structure [51] was used to classify parent residences as a major city, inner regional city, or outer city location. This classification was included as health disparities, including nutrition-related risk factors, exist between major city and inner regional cities (characterized by shorter distances by road to access services) and outer city locations (characterized by longer distances by road to services) [52].

A measure of between-group differences in primary and secondary outcomes was assessed for paired data using 2-level

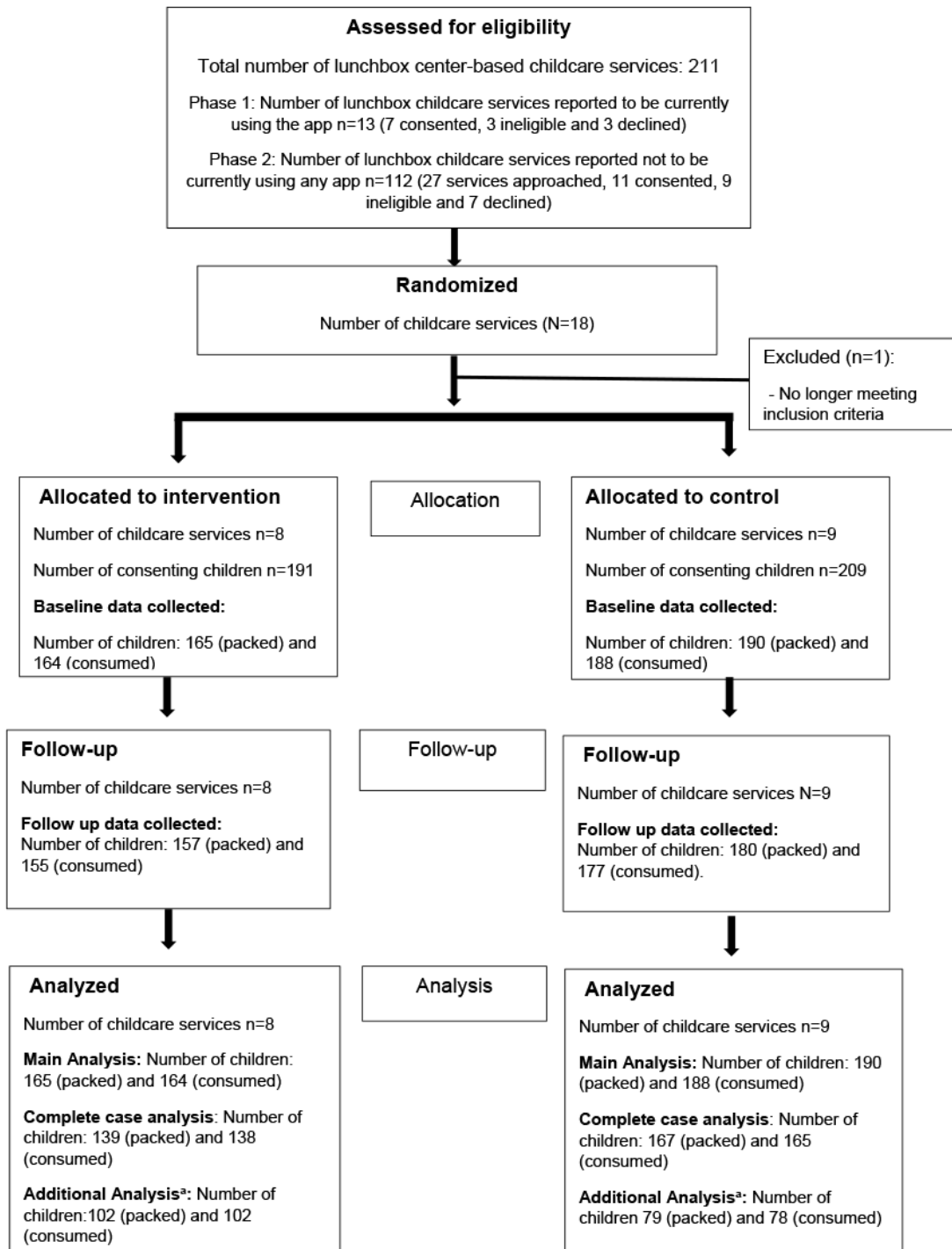
hierarchical linear regression models. Models were adjusted for potential ECEC service level clustering through a service random effect and controlled for baseline service EPAO score and if ECEC services were existing app users before the trial. An intention-to-treat analysis was undertaken [53], with multiple imputations undertaken for primary and secondary outcomes using the SAS MI and MIANALYZE procedure. Complete case analysis and prespecified subgroup analyses by child gender and SES were undertaken by adding a group by subgroup interaction fixed effect. SES was determined by parent postcodes and classified as being in the top or bottom 50% of NSW according to the Socioeconomic Indexes for Areas [54]. A sensitivity outcome analysis, modeled similarly to the complete case analysis, was undertaken whereby the effects of the intervention were assessed in parents known to have only downloaded the app (181/400, 51%). In addition to the analysis prespecified as part of the trial protocol [32], an exploratory analysis was performed to explore any association between how often the service routinely used the app (mean hours per week) and the parent viewing rate of within-app messages using Spearman rank order correlation.

Results

Sample

The combined consent rate of ECEC services was 45% (18/40; 40 services approached to reach the required 18 services). A service in the intervention arm ceased using the required app and withdrew from the study (before baseline data but after random allocation). Within the 17 consenting ECEC services, 400 parent-child dyads consented to participate in the baseline data collection. The mean parent consent rate was 51%. Data on packed lunch box contents were collected for 88.8% (355/400) of the children at baseline and for 84.3% (337/400) after the intervention. Consumption data were collected for 88% (352/400) of children at baseline and for 84.3% (337/400) after the intervention ([Figure 1](#)). The number of parents from the intervention group completing the web-based survey after the intervention was 21.6% (41/190). Parents who completed the survey were less likely to have a Technical and Further Education certificate or diploma ($P=.04$) or be from an inner regional city. ($P=.04$). Postintervention service pen and paper surveys were completed by all intervention ECEC services.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. ECEC: early childhood education and care.



^aAnalysis including only parents known to have downloaded the app required to receive the intervention

Parent and Service Characteristics

Baseline characteristics of ECEC services, parents, and children with lunch box data at baseline are reported in Table 1. Characteristics were similar across groups; however, the intervention group had a higher proportion of parents located in the outer regional areas than the control group. As the

intervention was delivered at the cluster level rather than at the individual level, geographical remoteness was accounted for by stratifying services by this factor during randomization and not controlled for as part of the analysis. Before the intervention, 12% (2/17) of ECEC services in the intervention group were existing users of the app, and 35% (6/17) of ECEC services had not been using any app.

Table 1. Characteristics of ECEC^a services (n=17) and children (n=355).

Characteristics	Intervention	Control
Service		
ECEC services, n (%)	8 (47)	9 (53)
ECEC service type		
Preschool	7 (88)	7 (78)
Long day care	1 (12)	2 (22)
Hours opened per day, mean (SD)	8.2 (1.54)	8.2 (0.88)
Number of child enrollments, mean (SD)	68.1 (27.4)	68.7 (21.58)
Modified EPAO ^b score (score out of 20), mean (SD) ^c	18.35 (2.33)	16.63 (2.86)
Parent and children dyads		
Dyads, n (%)	165 (46.5)	190 (53.5)
Child age (years) ^d , mean (SD)	3.9 (0.68)	3.9 (0.67)
Child gender, n (%)		
Female	86 (52.1)	98 (51.6)
Male	79 (47.9)	92 (48.4)
Parent SEIFA^e (based on postcode)^f, n (%)		
Most disadvantaged	98 (62.8)	102 (55.7)
Least disadvantaged	58 (37.2)	81 (44.3)
Parent education level^g, n (%)		
Attended or completed high school	37 (23.4)	40 (22.2)
Technical or Further Education (TAFE) certificate or diploma	50 (31.6)	59 (32.8)
Completed university or college degree or higher	71 (44.9)	81 (45)
ASGSRs^{h,i}, n (%)		
Major cities and inner regional	100 (64.1)	157 (87.2)
Outer regional	56 (35.9)	26 (14.4)

^aECEC: early childhood education and care.

^bEPAO: Environment and Policy Assessment Observation.

^cMissing data (control: n=1).

^dMissing data (intervention: n=15; control: n=10).

^eSEIFA: Socioeconomic Indexes for Areas. ^fMissing data (intervention: n=9; control n=7).

^gMissing data (intervention: n=7; control: n=10).

^hASGSRs: Australian Statistical Geography Standard Remoteness Structure.

ⁱMissing data (intervention: n=9; control: n=10).

Packing and Consumption of Foods From Lunch Boxes

The results for the primary outcomes after multiple imputation and for complete case analysis are presented in [Table 2](#). There was no significant difference between groups for the primary outcomes (kilojoules from discretionary foods and kilojoules from saturated fat, free sugars, and grams of sodium from all foods) packed or consumed. In addition, there was no difference in secondary outcomes, including packed or consumed serves

of any core food groups or discretionary food serves between groups ([Table 3](#)). Data regarding the nutrients and food groups consumed are presented in [Multimedia Appendix 5](#). The sensitivity analysis, which included only those parents known to have downloaded the app (181/355, 51%), also did not detect any statistically significant differences ([Multimedia Appendix 6](#)). The subgroup analysis by gender and SES demonstrated no significant differences between groups ([Multimedia Appendix 7](#)).

Table 2. Mean change in total energy, energy from discretionary foods, and associated nutrients by group (packed).

Energy and nutrients	Intervention (n=139)		Control (n=167)		Imputed difference postintervention ^a		Complete case analysis ^a	
	Baseline, mean (SD)	Postintervention, mean (SD)	Baseline, mean (SD)	Postintervention, mean (SD)	Mean difference (95% CI)	P value ^b	Mean difference (95% CI)	P value ^b
Packed								
Total energy (kJ)	2892.68 (850.71)	2820.13 (774.26)	2077.76 (672.89)	2762.47 (829.61)	39.53 (-264.57 to 343.64)	.80	19.45 (-293.05 to 331.94)	.90
Energy from discretionary foods (kJ)	774.58 (644.77)	799.26 (693.29)	800.78 (724.92)	712.03 (685.86)	77.84 (-163.49 to 319.18)	.53	62.78 (-190.87 to 316.42)	.60
Saturated fat (g)	9.64 (4.79)	8.62 (4.76)	8.97 (5.03)	7.84 (4.43)	0.72 (-0.52 to 1.96)	.26	0.56 (-1.00 to 2.12)	.45
Free sugars (g)	13.48 (9.65)	14.32 (11.94)	14.32 (11.94)	12.30 (11.59)	2.27 (-0.14 to 4.68)	.06	1.04 (-2.23 to 4.32)	.50
Sodium (mg)	1010.17 (426.86)	986.98 (365.59)	1026.12 (420.24)	934.74 (390.30)	49.93 (-68.86 to 168.72)	.41	60.78 (-95.85 to 217.42)	.42

^aAll data adjusted for baseline and clustering and service Environment and Policy Assessment Observation score at baseline.

^bStatistical significance inferred by P values <.01.

Table 3. Mean change in serves of discretionary foods and core food groups packed by group.

Food group	Intervention (n=139)		Control (n=167)		Imputed difference postintervention ^a		Complete case analysis ^b	
	Baseline, mean (SD)	Postintervention, mean (SD)	Baseline, mean (SD)	Postintervention, mean (SD)	Mean difference (95% CI)	P value ^c	Mean difference (95% CI)	P value ^c
Packed (serves)								
Discretionary foods ^d	1.29 (1.07)	1.33 (1.16)	1.33 (1.16)	1.19 (1.14)	0.13 (-0.27 to 0.53)	.53	0.10 (-0.32 to 0.53)	.60
Breads and cereals ^e	2.03 (1.02)	2.10 (1.01)	2.17 (0.96)	2.30 (1.06)	-0.27 (-0.58 to 0.05)	.09	-0.21 (-0.51 to 0.09)	.15
Fruit ^f	1.27 (0.85)	1.22 (0.82)	1.31 (0.89)	2.30 (1.06)	-0.10 (-0.47 to 0.27)	.61	-0.12 (-0.50 to 0.26)	.59
Vegetables ^g	0.25 (0.40)	0.20 (0.36)	0.21 (0.38)	0.21 (0.38)	0.00 (-0.10 to 0.10)	.97	-0.02 (-0.13 to 0.08)	.63
Dairy ^h	0.71 (0.52)	0.61 (0.49)	0.57 (0.50)	0.57 (0.50)	0.02 (-0.12 to 0.17)	.75	0.00 (-0.14 to 0.15)	.98
Meat and alternatives ^b	0.06 (0.21)	0.06 (0.20)	0.07 (0.23)	0.05 (0.16)	0.01 (-0.04 to 0.06)	.60	0.00 (-0.05 to 0.06)	.87

^aAll data adjusted for baseline and clustering and service Environment and Policy Assessment Observation score at baseline.

^bMeat and alternatives: examples of 1 serve=65 g of cooked lean meat, 80 g of cooked poultry, 100 g cooked fish, 2 large eggs, 1 cup of legumes or beans [6].

^cStatistical significance inferred by P values <.01.

^dCalculated using 600 kJ equivalents; that is, approximately 2 scoops of ice cream, 50 to 60 g of processed meats, 30 g of salty crackers, 2 to 3 sweet biscuits, and 1 (40 g) donut [6].

^eBreads and cereals: examples of 1 serve=1 slice of bread; half medium roll; half cup of cooked rice, pasta, or noodles; and two-third cup wheat cereal flakes [6].

^fVegetables: examples of 1 serve=half cup cooked vegetables; half cup beans, peas, or lentils; 1 cup of leafy green or raw vegetables; half medium potato; and 1 medium tomato [6].

^gFruit: examples of 1 serve=1 medium apple, 2 small fruits, 1 cup of diced or canned fruit, and 30 g of dried fruit [6].

^hDairy and alternatives: examples of 1 serve=1 cup milk, 2 slices of hard cheese (40 g), and three-fourth cup yogurt [6].

Process Evaluation

Intervention Fidelity

All (11/11, 100%) of the push notifications messages were delivered via the app to intervention ECEC services as planned. A message video link failed at week 9 and was resent as part of an unplanned additional message in week 11. All service managers reported sending the preintervention and midway planned support messages to parents (7/7, 100%; missing data 1/7, 14%).

App Downloads and SWAP IT Program Use

All ECEC services had access to the app, and it was available for all parents to download. The percentage of intervention families with data at baseline (as included in the multiple imputation analysis) and known to have the intervention app was 61.8% (102/165; missing data on app ownership 15/165, 9.1%).

Data on the number of unique within-app message views are presented in [Multimedia Appendix 4](#) [50]. As app analytics only provided deidentified data, data specific to parents in the trial could not be separated from data on all users of the app at the intervention ECEC services. The number of unique within-app message views decreased over time, with a mean of 139 (SD 42.7) views per message and a mean viewing rate of 26% (SD 14.9%).

SWAP IT Program Acceptability

Data on the acceptability of the *SWAP IT* program by service managers and parents are reported in [Multimedia Appendix 4](#) [50]. Parents who completed the survey reported that they liked the *SWAP IT* program (34/41, 83%) and found the program useful (33/41, 80%) and easy to use (36/41, 87%). Only 57% (4/7) of ECEC services rated the overall program as useful; however, most agreed that the resources within the program were helpful for families (6/7, 86%). Both ECEC services (8/8, 100%) and families that completed the survey (41/41, 100%) agreed or had no feelings either way regarding the timing and frequency of the push notification measures.

Feasibility of Ongoing Use of the App

Most ECEC services and parents who completed the survey agreed that it was appropriate to deliver lunch box information via the app (6/8, 75% and 40/50, 80%, respectively). During the intervention, most intervention ECEC services reported using the app for functions other than the delivery of the program (6/7, 86%; eg, for distribution of parent newsletters). Self-reported use of the app ranged from 0 to 2 hours per week. Only 57% (4/7) of ECEC services indicated that they planned to continue to use the app, with 29% (2/7) of ECEC services indicating they were unsure and 14% (1/7) of services reporting that they did not plan to continue with the app after the end of the *SWAP IT* program.

Cointervention and Adverse Events

The service reported changes in the frequency of parent complaints or concerns regarding healthy lunch box policy did not differ between the groups. No contamination was reported; that is, the app was not used to send any other health or nutrition

information. No ECEC services in the intervention or control groups reported exposure to additional nutrition interventions throughout the duration of the trial.

Outcomes From the Study Protocol Not Reported

Owing to the null findings, neither a cost-effectiveness analysis, as specified in the study protocol, nor the planned analysis of data collected on the usual daily dietary intake of children (to be able to detect any compensatory dietary behaviors) was conducted.

Association Between ECEC Service Hours of Routine Use of the App and Parent Viewing Rate

The exploratory analysis found an association between the number of routine hours of use of the app by ECEC services and the parent viewing rate of within-app messages; however, this association was not significant ($P=.21$; data not provided).

Discussion

Principal Findings

To our knowledge, the *SWAP IT Childcare* trial is the first trial to evaluate the efficacy of an mHealth intervention to improve lunch box contents in ECEC settings internationally. However, contrary to our hypothesis, the intervention had no impact on the primary outcome of the number of kilojoule from discretionary foods packed in children's lunch boxes. In addition, process evaluation found low parent app downloads and message viewing rates.

When looking to compare the findings of the trial with other published studies, we were unable to identify any other healthy eating-focused mHealth interventions or DHIs targeting parents and children through ECEC settings. Therefore, we have primarily compared the findings with those of the original *SWAP IT* trial publications conducted in the primary school setting [39].

Although the results of our parent survey data indicated that delivery of a lunch box program via the app was considered appropriate, the mean viewing rate of the within-app message was only 26% (SD 14.9). Furthermore, only 61.8% (102/165) of parents in the intervention group were known to have downloaded the app, leaving 38.2% (63/165) of parents unexposed to program content. Such process data are in contrast to findings reported in a *SWAP IT* pilot study undertaken in a primary (elementary) school setting [39]. The intervention in the *SWAP IT* pilot study was similar to this study, including push notifications and within-app message content; however, it also included the distribution of resources to parents, policy, and classroom resources [39]. The pilot study reported higher parent app downloads (89%) and message views (between 35% and 120% over the 10-week intervention) and was able to demonstrate a reduction in the packing of kilojoules from discretionary foods (-221 kJ; $P=.08$, nonsignificant) [39].

Contextual differences between the ECEC and school settings in the study region may potentially explain the differences in the findings between these trials. In the school trial, all schools were existing users of the app used to deliver the *SWAP IT* program compared with just 2 ECEC services in the intervention

arm of this trial. In addition, most parents were existing users of the app (355/400, 88.8%) [39]. Greater integration and preprogram use of the app may have meant that parents in the school trial were more likely to be routinely accessing the app to receive information and, in doing so, were more likely to access the *SWAP IT* program content. This premise is supported by our exploratory analysis, which found a positive nonsignificant association between how much time the ECEC services used the app for general communication with parents and the *SWAP IT* parent message viewing rate. In addition, it is possible that the inclusion of supportive ECEC service-based strategies targeting nutrition curriculum and policy, similar to that undertaken to the trial in schools, may have been useful in increasing the impact of the intervention.

Our ability to detect an effect on discretionary food packing may have also been limited by the lower than anticipated baseline packing of discretionary foods in the participating ECEC services [48]. The trial baseline data found ECEC child lunch boxes contained only a mean of 1.33 (SD 1.16) serves of discretionary foods [48]. This was considerably lower than the mean of 2.5 serves per lunch box reported in schools in the same region [55]. Although the serves of discretionary foods packed in the ECEC services exceeded recommended serves per day for our age group (ie, 0-0.5 serves) [6], the lower number of serves packed at baseline possibly limited scope for further improvement without a more intensive intervention approach. In contrast, a large scope for improving vegetable packing and consumption in children attending ECEC services was evident from the trial baseline data (only 0.3 mean vegetable serves packed and 0.1 serves consumed) [48]. As such, future iterations of the intervention would benefit from the inclusion of specific strategies to improve vegetable packing and intake, along with continued efforts to minimize discretionary foods.

The intervention's lack of impact on parent packing of discretionary foods may have also been because of the decline in the rate of program message opening over the course of the trial. The pattern of the continued drop-off in use observed in our study is similar to those reported in many mHealth and DHI interventions [39,56-59]. The drop-off in engagement occurred in the trial despite the inclusion of several features explicitly designed to enhance engagement (user consultation, use of behavior change techniques, reminders, and information being provided from a credible source) [60,61] and the addition of a specific intervention strategy to prompt ongoing use (service manager delivered support messages). Additional factors found to be associated with sustained DHI use in the literature include

perceived ongoing relevance, novelty value, tailoring, and self-monitoring [60,62]. Further qualitative exploration of parents' reasons for low engagement and consideration of additional evidence-based features that could be added to the program to enhance engagement would be beneficial.

Strengths and Limitations

The strengths of this study include explicit mapping of behavior change theory to design the intervention, use of objective gold standard measures to assess lunch box contents, high follow-up rate, comprehensive process measures, and randomized trial design. A key limitation of the trial is that the app analytics could only capture use data at the service level, meaning that the true proportion of parents who accessed the messages is unknown (as message views may have been from parents from the service but not participating in the evaluation component of the trial). This limited our ability to conclusively determine whether the null findings were generally more likely to be a result of poor access, the program content itself, or a combination. This limitation extends to our additional analysis of only those parents who were known to have downloaded the app. Finally, given the low survey response rate (41/190, 21.6%) and differences in characteristics in survey completers compared with the overall parent sample, our acceptability survey data may have been influenced by response bias. The generalizability of the findings to ECEC services outside of the study region may be limited.

Conclusions

The intervention failed to decrease energy from discretionary foods in children's lunch boxes. However, the use of apps was rated as an appropriate modality for delivering information related to the packing of healthier foods in lunch boxes. Process data suggest that the lack of impact may have been because of factors associated with the implementation of the intervention, such as the low parent uptake and use of the app, as well as the lower than expected levels of packing of discretionary foods at baseline. Given these limitations, the current feasibility of using mHealth interventions to target parent packing of healthier foods in lunch boxes is still uncertain. Future trials should invest time in collecting formative data on how ECEC services and parents currently use and engage with different digital platforms to best identify the ideal technology to deliver parent-focused nutrition interventions in this setting. In addition, given drop-offs in use over time, exploration of parent reasons behind poor ongoing use will also assist in enhancing program design and content to improve engagement.

Acknowledgments

The authors would like to acknowledge the staff involved in collecting the data and data entry, including Jayde Kerr, Monique Santarelli, and consulting statistician Christophe Lecathelinais. The authors would also like to extend their appreciation to the app provider Skoolbag and participating early childhood education and care services.

Conflicts of Interest

None declared.

Multimedia Appendix 1

CONSORT-EHEALTH checklist V 1.6.1.

[\[PDF File \(Adobe PDF File\), 2581 KB - jmir_v24i3e27760_app1.pdf \]](#)

Multimedia Appendix 2

Use of behaviour change techniques.

[\[DOCX File , 14 KB - jmir_v24i3e27760_app2.docx \]](#)

Multimedia Appendix 3

Adaptation process.

[\[DOCX File , 14 KB - jmir_v24i3e27760_app3.docx \]](#)

Multimedia Appendix 4

Process evaluation.

[\[DOCX File , 22 KB - jmir_v24i3e27760_app4.docx \]](#)

Multimedia Appendix 5

Secondary outcome data.

[\[DOCX File , 16 KB - jmir_v24i3e27760_app5.docx \]](#)

Multimedia Appendix 6

Sensitivity analysis.

[\[DOCX File , 14 KB - jmir_v24i3e27760_app6.docx \]](#)

Multimedia Appendix 7

Subgroup analysis.

[\[DOCX File , 17 KB - jmir_v24i3e27760_app7.docx \]](#)**References**

1. World Health Organization. Diet, nutrition and the prevention of chronic disease: report of a joint WHO/FAO expert consultation. In: Report of a Joint WHO/FAO Expert Consultation. Geneva: World Health Organisation; 2003.
2. GBD 2017 Diet Collaborators. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2019 May 11;393(10184):1958-1972 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(19\)30041-8](https://doi.org/10.1016/S0140-6736(19)30041-8)] [Medline: [30954305](https://pubmed.ncbi.nlm.nih.gov/30954305/)]
3. Drewnowski A, Rehm CD. Consumption of added sugars among US children and adults by food purchase location and food source. *Am J Clin Nutr* 2014 Sep;100(3):901-907 [[FREE Full text](#)] [doi: [10.3945/ajcn.114.089458](https://doi.org/10.3945/ajcn.114.089458)] [Medline: [25030785](https://pubmed.ncbi.nlm.nih.gov/25030785/)]
4. National diet and nutrition survey: results from years 5-6 (combined) of the rolling programme (2012/13 and 2013/14). Public Health England. 2012. URL: <https://www.gov.uk/government/statistics/ndns-results-from-years-5-and-6-combined> [accessed 2019-05-01]
5. Australian Institute of Health and Welfare. Nutrition Across the Life Stages. Cat. no. PHE 227. Canberra: AIHW; 2018.
6. National Health and Medical Research Council. Educator Guide. Canberra: National Health and Medical Research Council; 2013.
7. National Health and Medical Research Council. A review of the evidence to address targeted questions to inform the revision of the Australian Dietary Guidelines. In: Eat for Health. Canberra: Commonwealth of Australia; 2011.
8. 2015-2020 Dietary guidelines for Americans. 8th Edition. U.S. Department of Health and Human Services and U.S. Department of Agriculture. 2015 Dec. URL: https://health.gov/sites/default/files/2019-09/2015-2020_Dietary_Guidelines.pdf [accessed 2019-05-01]
9. The eat well guide. National Health Service. URL: <https://www.gov.uk/government/publications/the-eatwell-guide> [accessed 2019-05-01]
10. Canada's dietary guidelines - Canada's food guide. Government Of Canada. 2019 Jan. URL: <https://food-guide.canada.ca/en/guidelines/> [accessed 2019-05-01]
11. Larson N, Ward DS, Neelon SB, Story M. What role can child-care settings play in obesity prevention? A review of the evidence and call for research efforts. *J Am Diet Assoc* 2011 Sep;111(9):1343-1362. [doi: [10.1016/j.jada.2011.06.007](https://doi.org/10.1016/j.jada.2011.06.007)] [Medline: [21872698](https://pubmed.ncbi.nlm.nih.gov/21872698/)]
12. Mikkelsen MV, Husby S, Skov LR, Perez-Cueto FJ. A systematic review of types of healthy eating interventions in preschools. *Nutr J* 2014 Jun 06;13:56 [[FREE Full text](#)] [doi: [10.1186/1475-2891-13-56](https://doi.org/10.1186/1475-2891-13-56)] [Medline: [24906305](https://pubmed.ncbi.nlm.nih.gov/24906305/)]

13. Lucas PJ, Patterson E, Sacks G, Billich N, Evans CE. Preschool and school meal policies: an overview of what we know about regulation, implementation, and impact on diet in the UK, Sweden, and Australia. *Nutrients* 2017 Jul 11;9(7):736 [FREE Full text] [doi: [10.3390/nu9070736](https://doi.org/10.3390/nu9070736)] [Medline: [28696403](https://pubmed.ncbi.nlm.nih.gov/28696403/)]
14. Sweitzer SJ, Briley ME, Robert-Gray C. Do sack lunches provided by parents meet the nutritional needs of young children who attend child care? *J Am Diet Assoc* 2009 Jan;109(1):141-144. [doi: [10.1016/j.jada.2008.10.010](https://doi.org/10.1016/j.jada.2008.10.010)] [Medline: [19103336](https://pubmed.ncbi.nlm.nih.gov/19103336/)]
15. Child care and early childhood education in Australia. Australian Government Australian Institute of Family Studies. 2015. URL: <https://aifs.gov.au/publications/child-care-and-early-childhood-education-australia> [accessed 2018-09-20]
16. Healthy children initiative. Population Health Information Management System (PHIMS). URL: <https://www.health.nsw.gov.au/epidemiology/Pages/population-health-intervention.aspx> [accessed 2018-08-14]
17. Zask A, Adams JK, Brooks LO, Hughes DF. Tooty Fruity Vegie: an obesity prevention intervention evaluation in Australian preschools. *Health Promot J Austr* 2012 Apr;23(1):10-15. [doi: [10.1071/he12010](https://doi.org/10.1071/he12010)] [Medline: [22730932](https://pubmed.ncbi.nlm.nih.gov/22730932/)]
18. Kelly B, Hardy LL, Howlett S, King L, Farrell L, Hattersley L. Opening up Australian preschoolers' lunchboxes. *Aust N Z J Public Health* 2010 Jun;34(3):288-292. [doi: [10.1111/j.1753-6405.2010.00528.x](https://doi.org/10.1111/j.1753-6405.2010.00528.x)] [Medline: [20618271](https://pubmed.ncbi.nlm.nih.gov/20618271/)]
19. Roberts-Gray C, Briley ME, Ranjit N, Byrd-Williams CE, Sweitzer SJ, Sharma SV, et al. Efficacy of the Lunch is in the Bag intervention to increase parents' packing of healthy bag lunches for young children: a cluster-randomized trial in early care and education centers. *Int J Behav Nutr Phys Act* 2016 Jan 08;13:3 [FREE Full text] [doi: [10.1186/s12966-015-0326-x](https://doi.org/10.1186/s12966-015-0326-x)] [Medline: [26746876](https://pubmed.ncbi.nlm.nih.gov/26746876/)]
20. Hardy LL, King L, Kelly B, Farrell L, Howlett S. Munch and move: evaluation of a preschool healthy eating and movement skill program. *Int J Behav Nutr Phys Act* 2010;7:80 [FREE Full text] [doi: [10.1186/1479-5868-7-80](https://doi.org/10.1186/1479-5868-7-80)] [Medline: [21047434](https://pubmed.ncbi.nlm.nih.gov/21047434/)]
21. Roberts-Gray C, Sweitzer SJ, Ranjit N, Potratz C, Rood M, Romo-Palafox MJ, et al. Structuring process evaluation to forecast use and sustainability of an intervention: theory and data from the efficacy trial for lunch is in the bag. *Health Educ Behav* 2017 Aug;44(4):559-569. [doi: [10.1177/1090198116676470](https://doi.org/10.1177/1090198116676470)] [Medline: [27864471](https://pubmed.ncbi.nlm.nih.gov/27864471/)]
22. McCrabb S, Lane C, Hall A, Milat A, Bauman A, Sutherland R, et al. Scaling-up evidence-based obesity interventions: a systematic review assessing intervention adaptations and effectiveness and quantifying the scale-up penalty. *Obes Rev* 2019 Jul;20(7):964-982. [doi: [10.1111/obr.12845](https://doi.org/10.1111/obr.12845)] [Medline: [30868745](https://pubmed.ncbi.nlm.nih.gov/30868745/)]
23. Dev DA, Byrd-Williams C, Ramsay S, McBride B, Srivastava D, Murriel A, et al. Engaging parents to promote children's nutrition and health. *Am J Health Promot* 2017 Mar;31(2):153-162. [doi: [10.1177/0890117116685426](https://doi.org/10.1177/0890117116685426)] [Medline: [28423928](https://pubmed.ncbi.nlm.nih.gov/28423928/)]
24. Nathan N, Janssen L, Sutherland R, Hodder RK, Evans CE, Booth D, et al. The effectiveness of lunchbox interventions on improving the foods and beverages packed and consumed by children at centre-based care or school: a systematic review and meta-analysis. *Int J Behav Nutr Phys Act* 2019 Apr 29;16(1):38 [FREE Full text] [doi: [10.1186/s12966-019-0798-1](https://doi.org/10.1186/s12966-019-0798-1)] [Medline: [31036038](https://pubmed.ncbi.nlm.nih.gov/31036038/)]
25. Zarnowiecki D, Mauch CE, Middleton G, Matwiejczyk L, Watson WL, Dibbs J, et al. A systematic evaluation of digital nutrition promotion websites and apps for supporting parents to influence children's nutrition. *Int J Behav Nutr Phys Act* 2020 Feb 10;17(1):17 [FREE Full text] [doi: [10.1186/s12966-020-0915-1](https://doi.org/10.1186/s12966-020-0915-1)] [Medline: [32041640](https://pubmed.ncbi.nlm.nih.gov/32041640/)]
26. Hingle M, Patrick H. There are thousands of apps for that: navigating mobile technology for nutrition education and behavior. *J Nutr Educ Behav* 2016 Mar;48(3):213-218. [doi: [10.1016/j.jneb.2015.12.009](https://doi.org/10.1016/j.jneb.2015.12.009)] [Medline: [26965099](https://pubmed.ncbi.nlm.nih.gov/26965099/)]
27. Curtis KE, Lahiri S, Brown KE. Targeting parents for childhood weight management: development of a theory-driven and user-centered healthy eating app. *JMIR Mhealth Uhealth* 2015;3(2):e69 [FREE Full text] [doi: [10.2196/mhealth.3857](https://doi.org/10.2196/mhealth.3857)] [Medline: [26088692](https://pubmed.ncbi.nlm.nih.gov/26088692/)]
28. Rose T, Barker M, Maria JC, 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](https://doi.org/10.1016/j.jadohealth.2017.05.024)] [Medline: [28822682](https://pubmed.ncbi.nlm.nih.gov/28822682/)]
29. Tully L, Burls A, Sorensen J, El-Moslemany R, O'Malley G. Mobile health for pediatric weight management: systematic scoping review. *JMIR Mhealth Uhealth* 2020 Jun 03;8(6):e16214 [FREE Full text] [doi: [10.2196/16214](https://doi.org/10.2196/16214)] [Medline: [32490849](https://pubmed.ncbi.nlm.nih.gov/32490849/)]
30. Røed M, Medin AC, Vik FN, Hillesund ER, Van Lippevelde W, Campbell K, et al. Effect of a parent-focused eHealth intervention on children's fruit, vegetable, and discretionary food intake (Food4toddlers): randomized controlled trial. *J Med Internet Res* 2021 Feb 16;23(2):e18311 [FREE Full text] [doi: [10.2196/18311](https://doi.org/10.2196/18311)] [Medline: [33591279](https://pubmed.ncbi.nlm.nih.gov/33591279/)]
31. 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](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]
32. Pond N, Finch M, Sutherland R, Wolfenden L, Nathan N, Kingsland M, et al. Cluster randomised controlled trial of an m-health intervention in centre-based childcare services to reduce the packing of discretionary foods in children's lunchboxes: study protocol for the 'SWAP IT Childcare' trial. *BMJ Open* 2019 Jun 01;9(5):e026829 [FREE Full text] [doi: [10.1136/bmjopen-2018-026829](https://doi.org/10.1136/bmjopen-2018-026829)] [Medline: [31154306](https://pubmed.ncbi.nlm.nih.gov/31154306/)]
33. Childhood education and care. Australian Bureau of Statistics. 2017. URL: <http://www.abs.gov.au/ausstats/abs@.nsf/mf/4402.0> [accessed 2020-02-01]
34. New South Wales Health website, Hunter New England. New South Wales Health. URL: <https://www.health.nsw.gov.au/lhd/pages/hnelhd.aspx> [accessed 2020-10-01]

35. Census of population health and housing-community profiles. Australian Bureau of Statistics. 2016. URL: <http://www.abs.gov.au/websitedbs/D3310114.nsf/Home/2016%20Census%20Community%20Profiles> [accessed 2018-05-01]
36. Burns CM, Gibbon P, Boak R, Baudinette S, Dunbar JA. Food cost and availability in a rural setting in Australia. *Rural Remote Health* 2004;4(4):311 [FREE Full text] [Medline: 15887990]
37. Ward PR, Coveney J, Verity F, Carter P, Schilling M. Cost and affordability of healthy food in rural South Australia. *Rural Remote Health* 2012;12:1938 [FREE Full text] [Medline: 22533350]
38. Kang M, Ragan BG, Park J. Issues in outcomes research: an overview of randomization techniques for clinical trials. *J Athl Train* 2008;43(2):215-221 [FREE Full text] [doi: 10.4085/1062-6050-43.2.215] [Medline: 18345348]
39. Sutherland R, Nathan N, Brown A, Yoong S, Finch M, Lecathelinais C, et al. A randomized controlled trial to assess the potential efficacy, feasibility and acceptability of an m-health intervention targeting parents of school aged children to improve the nutritional quality of foods packed in the lunchbox 'SWAP IT'. *Int J Behav Nutr Phys Act* 2019 Jul 02;16(1):54 [FREE Full text] [doi: 10.1186/s12966-019-0812-7] [Medline: 31266506]
40. Sutherland R, Brown A, Nathan N, Yoong S, Janssen L, Chooi A, et al. A multicomponent mHealth-based intervention (SWAP IT) to decrease the consumption of discretionary foods packed in school lunchboxes: type I effectiveness-implementation hybrid cluster randomized controlled trial. *J Med Internet Res* 2021 Jun 24;23(6):e25256 [FREE Full text] [doi: 10.2196/25256] [Medline: 34185013]
41. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. *Implement Sci* 2011;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
42. Wiltsey Stirman S, Baumann AA, Miller CJ. The FRAME: an expanded framework for reporting adaptations and modifications to evidence-based interventions. *Implement Sci* 2019 Jun 06;14(1):58 [FREE Full text] [doi: 10.1186/s13012-019-0898-y] [Medline: 31171014]
43. Rudd BN, Davis M, Beidas RS. Integrating implementation science in clinical research to maximize public health impact: a call for the reporting and alignment of implementation strategy use with implementation outcomes in clinical research. *Implement Sci* 2020 Nov 25;15(1):103 [FREE Full text] [doi: 10.1186/s13012-020-01060-5] [Medline: 33239097]
44. Burrows TL, Martin RJ, Collins CE. A systematic review of the validity of dietary assessment methods in children when compared with the method of doubly labeled water. *J Am Diet Assoc* 2010 Oct;110(10):1501-1510. [doi: 10.1016/j.jada.2010.07.008] [Medline: 20869489]
45. Liz Martins M, Cunha LM, Rodrigues SS, Rocha A. Determination of plate waste in primary school lunches by weighing and visual estimation methods: a validation study. *Waste Manag* 2014 Aug;34(8):1362-1368. [doi: 10.1016/j.wasman.2014.03.020] [Medline: 24841068]
46. Foodworks. Version 9 ed. Xyris Software. 2017. URL: <https://xyris.com.au/> [accessed 2018-01-10]
47. Australian health survey: user's guide, 2011-13. Australian Bureau of Statistics. 2013. URL: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/4363.0.55.0012011-13?OpenDocument> [accessed 2018-06-01]
48. Pearson N, Wolfenden L, Finch M, Yoong SL, Kingsland M, Nathan N, et al. A cross-sectional study of packed lunchbox foods and their consumption by children in early childhood education and care services. *Nutr Diet* 2021 Sep;78(4):397-405. [doi: 10.1111/1747-0080.12632] [Medline: 32830413]
49. Ward D, Hales D, Haverly K, Marks J, Benjamin S, Ball S, et al. An instrument to assess the obesogenic environment of child care centers. *Am J Health Behav* 2008;32(4):380-386. [doi: 10.5555/ajhb.2008.32.4.380] [Medline: 18092898]
50. Weiner BJ, Lewis CC, Stanick C, Powell BJ, Dorsey CN, Clary AS, et al. Psychometric assessment of three newly developed implementation outcome measures. *Implement Sci* 2017 Dec 29;12(1):108 [FREE Full text] [doi: 10.1186/s13012-017-0635-3] [Medline: 28851459]
51. 1270.0.55.005 - Australian Statistical Geography Standard (ASGS): volume 5 - Remoteness structure. Australian Bureau of Statistics. 2016 Jul. URL: <https://www.abs.gov.au/ausstats/abs@.nsf/PrimaryMainFeatures/1270.0.55.005?OpenDocument> [accessed 2018-06-01]
52. Rural and Remote Health. Australian Institute of Health and Welfare. 2020. URL: <https://www.aihw.gov.au/reports/australias-health/rural-and-remote-health> [accessed 2021-06-01]
53. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention to treat analysis in randomised trials with missing outcome data. *Br Med J* 2011 Feb 07;342:d40 [FREE Full text] [Medline: 21300711]
54. Information paper: an introduction to Socio-Economic Indexes for Areas (SEIFA). Australian Bureau of Statistics. 2006. URL: <https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/2039.0Main+Features12006?OpenDocument> [accessed 2020-02-05]
55. Sutherland R, Nathan N, Brown A, Yoong S, Reynolds R, Walton A, et al. A cross-sectional study to determine the energy density and nutritional quality of primary-school children's lunchboxes. *Public Health Nutr* 2020 Apr;23(6):1108-1116. [doi: 10.1017/S1368980019003379] [Medline: 31969199]
56. 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]
57. Vaghefi I, Tulu B. The continued use of mobile health apps: insights from a longitudinal study. *JMIR Mhealth Uhealth* 2019 Aug 29;7(8):e12983 [FREE Full text] [doi: 10.2196/12983] [Medline: 31469081]

58. Hammersley ML, Okely AD, Batterham MJ, Jones RA. An internet-based childhood obesity prevention program (Time2bHealthy) for parents of preschool-aged children: randomized controlled trial. *J Med Internet Res* 2019 Feb 08;21(2):e11964 [FREE Full text] [doi: [10.2196/11964](https://doi.org/10.2196/11964)] [Medline: [30735139](https://pubmed.ncbi.nlm.nih.gov/30735139/)]
59. Russell CG, Denney-Wilson E, Laws RA, Abbott G, Zheng M, Lymer SJ, et al. Impact of the growing healthy mHealth program on maternal feeding practices, infant food preferences, and satiety responsiveness: quasi-experimental study. *JMIR Mhealth Uhealth* 2018 Apr 25;6(4):e77 [FREE Full text] [doi: [10.2196/mhealth.9303](https://doi.org/10.2196/mhealth.9303)] [Medline: [29695373](https://pubmed.ncbi.nlm.nih.gov/29695373/)]
60. Short C, Rebar A, Plotnikoff R, Vandelanotte C. Designing engaging online behaviour change interventions: a proposed model of user engagement. *Eur Health Psychol* 2015;17(1):32-38 [FREE Full text]
61. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. *Transl Behav Med* 2017 Jun;7(2):254-267 [FREE Full text] [doi: [10.1007/s13142-016-0453-1](https://doi.org/10.1007/s13142-016-0453-1)] [Medline: [27966189](https://pubmed.ncbi.nlm.nih.gov/27966189/)]
62. Taki S, Lymer S, Russell CG, Campbell K, Laws R, Ong K, et al. Assessing user engagement of an mHealth intervention: development and implementation of the growing healthy app engagement index. *JMIR Mhealth Uhealth* 2017 Jun 29;5(6):e89 [FREE Full text] [doi: [10.2196/mhealth.7236](https://doi.org/10.2196/mhealth.7236)] [Medline: [28663164](https://pubmed.ncbi.nlm.nih.gov/28663164/)]

Abbreviations

BCW: behavior change wheel

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and online Telehealth

DHI: digital health intervention

ECEC: early childhood education and care

EPAO: Environment and Policy Assessment Observation

HNE: Hunter New England

mHealth: mobile health

NSW: New South Wales

SES: socioeconomic status

Edited by R Kukafka; submitted 06.02.21; peer-reviewed by T Swindle, D Dev; comments to author 27.03.21; revised version received 22.07.21; accepted 17.12.21; published 17.03.22.

Please cite as:

Pearson N, Finch M, Sutherland R, Kingsland M, Wolfenden L, Wedesweiler T, Herrmann V, Yoong SL

An mHealth Intervention to Reduce the Packing of Discretionary Foods in Children's Lunch Boxes in Early Childhood Education and Care Services: Cluster Randomized Controlled Trial

J Med Internet Res 2022;24(3):e27760

URL: <https://www.jmir.org/2022/3/e27760>

doi: [10.2196/27760](https://doi.org/10.2196/27760)

PMID: [35297768](https://pubmed.ncbi.nlm.nih.gov/35297768/)

©Nicole Pearson, Meghan Finch, Rachel Sutherland, Melanie Kingsland, Luke Wolfenden, Taya Wedesweiler, Vanessa Herrmann, Sze Lin Yoong. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 17.03.2022. 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.

Original Paper

Education on Depression in Mental Health Apps: Systematic Assessment of Characteristics and Adherence to Evidence-Based Guidelines

Laura Martinengo¹, MD; Anne-Claire Stona¹, MD; Lorainne Tudor Car^{2,3}, MSc, MD, PhD; Jimmy Lee^{4,5}, MBBS, MMed, MCI; Konstadina Griva¹, PhD; Josip Car^{1,3}, MD, PhD

¹Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore

²Family Medicine and Primary Care, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore

³Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

⁴Department of Psychosis & North Region, Institute of Mental Health, Singapore, Singapore

⁵Neuroscience and Mental Health, Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore

Corresponding Author:

Josip Car, MD, PhD

Centre for Population Health Sciences

Lee Kong Chian School of Medicine

Nanyang Technological University Singapore

11 Mandalay Rd

Clinical Sciences Building Level 18

Singapore, 308232

Singapore

Phone: 65 69047017

Email: josip.car@ntu.edu.sg

Abstract

Background: Suboptimal understanding of depression and mental health disorders by the general population is an important contributor to the wide treatment gap in depression. Mental health literacy encompasses knowledge and beliefs about mental disorders and supports their recognition, management, and prevention. Besides knowledge improvement, psychoeducational interventions reduce symptoms of depression, enhance help-seeking behavior, and decrease stigma. Mental health apps often offer educational content, but the trustworthiness of the included information is unclear.

Objective: The aim of this study is to systematically evaluate adherence to clinical guidelines on depression of the information offered by mental health apps available in major commercial app stores.

Methods: A systematic assessment of the educational content regarding depression in the apps available in the Apple App Store and Google Play was conducted in July 2020. A systematic search for apps published or updated since January 2019 was performed using 42matters. Apps meeting the inclusion criteria were downloaded and assessed using two smartphones: an iPhone 7 (iOS version 14.0.1) and a Sony XPERIA XZs (Android version 8.0.0). The 156-question assessment checklist comprised general characteristics of apps, appraisal of 38 educational topics and their adherence to evidence-based clinical guidelines, as well as technical aspects and quality assurance. The results were tabulated and reported as a narrative review, using descriptive statistics.

Results: The app search retrieved 2218 apps, of which 58 were included in the analysis (Android apps: n=29, 50%; iOS apps: n=29, 50%). Of the 58 included apps, 37 (64%) apps offered educational content within a more comprehensive depression or mental health management app. Moreover, 21% (12/58) of apps provided non-evidence-based information. Furthermore, 88% (51/58) of apps included up to 20 of the educational topics, the common ones being listing the symptoms of depression (52/58, 90%) and available treatments (48/58, 83%), particularly psychotherapy. Depression-associated stigma was mentioned by 38% (22/58) of the apps, whereas suicide risk was mentioned by 71% (41/58), generally as an item in a list of symptoms. Of the 58 included apps, 44 (76%) highlighted the importance of help seeking, 29 (50%) emphasized the importance of involving the user's support network. In addition, 52% (30/58) of apps referenced their content, and 17% (10/58) included advertisements.

Conclusions: Information in mental health and depression apps is often brief and incomplete, with 1 in 5 apps providing non-evidence-based information. Given the unmet needs and stigma associated with the disease, it is imperative that apps seize

the opportunity to offer quality, evidence-based education or point the users to relevant resources. A multistakeholder consensus on a more stringent development and publication process for mental health apps is essential.

(*J Med Internet Res* 2022;24(3):e28942) doi:[10.2196/28942](https://doi.org/10.2196/28942)

KEYWORDS

health literacy; mental health literacy; depression; mobile apps; apps; telemedicine; mHealth; self-management; mobile phone

Introduction

Background

Depression affects >264 million people worldwide and was the third major contributor of years lost to disability in 2017 [1]. Yet, approximately 50% of the people living in high-income countries and at least 75% of the people living in low- and middle-income countries [2,3] have neither been diagnosed nor receive treatment. Untreated depression is associated with increased morbidity and mortality, including death by suicide, poverty and unemployment at an individual level, and significant increase in health expenditure at health system level because of increased health care use and decreased workforce productivity of the affected individuals [3,4]. Although the reasons for this wide treatment gap are multifarious, suboptimal understanding of depression and mental health disorders in the general population is considered an important contributor to low use of mental health services. Studies assessing the general population's mental health literacy through the use of standardized clinical vignettes have shown poor recognition of common mental health disorders [5-7].

Mental Health Literacy

Mental health literacy is defined as “the knowledge and beliefs about mental disorders, which aid their recognition, management or prevention” [8]. It encompasses the acquisition of factual knowledge of mental health disorders as well as development of competencies and beliefs enabling prevention, early recognition of symptoms, help seeking, self-help, and provision of first aid to others [5]. Evidence-based clinical guidelines for management of depression indicate that psychoeducation is pivotal in terms of disease management [9-12] and necessary to expand the shared-decision model of care in mental health services [13].

Educational Topics

Educational topics focused on depression may include information related to its natural history, symptoms and signs, treatment options such as pharmacotherapy and a range of nonpharmacological treatments and possible side effects, as well as information related to prognosis and effective self-management interventions [9]. Psychoeducational interventions have been shown to increase mental health literacy and, most importantly, to yield small but significant reductions in symptoms of depression and mental distress [14,15], hence offering simple, inexpensive, and readily available tools for symptom management [14].

There are other notable benefits of psychoeducation, particularly stigma reduction [16-18] and increased help-seeking behavior [6,19-21]. Mental health stigma remains a significant barrier to

the use of mental health services because it affects the access and quality of health care provision for people living with mental health disorders and depression [22,23]. Social stigma or public stigma are linked to discrimination, avoidance, and inadequate treatment [24], whereas self-stigma or internalized stigma may erode self-esteem and self-efficacy or lead to anger or indifference [24], which may in turn hamper help seeking [25]. Although multilevel coordinated and sustained efforts are needed to mitigate the prevailing mental health stigma, simple interventions such as the use of personal narratives have shown promise in fostering depression awareness [26,27] and more positive attitudes among the public [26,28,29].

Mental Health Apps

Since the early 2000s, the search for health information has shifted increasingly to the web [30], and this trend has further intensified with the advent of smartphones [31] and mobile apps. A recent systematic review on the provision of medical education using smartphones reported that approximately two-thirds of the reviewed interventions were effective in improving patients' knowledge and clinical outcomes [32]. The Apple App Store and Google Play, which are the major sources of apps worldwide [33], currently include >10,000 mental health apps [34] that offer a wide range of functionalities comprising education, screening, and self-management programs for a wide range of disorders [35,36]. However, most of the published apps have not been assessed in clinical trials and are not evidence based [34,37].

Given the increasing popularity of health apps, a number of studies evaluating direct-to-consumer apps have been published in recent years, including the general characteristics of the apps [38], key features and functionalities of highly downloaded [39] or highly rated apps [40], techniques commonly used in face-to-face psychotherapy [41], or adherence of self-guided app interventions to evidence-based clinical guidelines [42]. These studies include a variety of assessment methodologies, from researcher-developed checklists [38,41,42] used in specific projects to the use of standardized assessment tools such as the Mobile App Rating Scale [36,43] or the Organization for the Review of Care and Health Applications–24 Question Assessment [44], focusing on the usability and technical aspects of the apps. In contrast, the health-related content of apps is seldom included in assessment checklists.

Existing mental health apps often include educational components, either as its sole functionality or within a depression or mental health management app. The trustworthiness, adherence to evidence, and depth of information offered by these apps is unclear because, to date, no reviews offering an in-depth assessment of the content of education on

depression modules provided by mental health apps has been published.

Therefore, this study aims to systematically evaluate the adherence to evidence-based clinical guidelines of the information on depression provided by mental health and depression apps available in the Apple App Store and Google Play.

Methods

We followed a rigorous assessment process, developed at our center and used in previous app assessment projects [45-48], by adapting systematic review methodology for the app search, selection, assessment, and data analysis.

Development of the Assessment Criteria

The research team designed the assessment criteria to evaluate the clinical and technical features of apps. The criteria included the following three domains:

1. *General attributes*, as described in the app store description, including developer, category, ratings, education-delivery format, target group, country of origin, and cost.
2. *Appraisal of depression education modules*, comprising the scope of information provided by the app and its adherence to evidence-based clinical guidelines from the Royal Australian and New Zealand College of Psychiatrists [11], Singapore's Ministry of Health [49], the United Kingdom's National Institute for Health and Care Excellence [50], and the American Psychiatric Association [9], as well as concordance with expert opinions on content of mental health awareness public campaigns [51,52]. The criteria included 46 questions encompassing 38 educational topics on symptoms and diagnostic criteria of depression, natural

history, importance of help seeking, treatment (including pharmacological and nonpharmacological treatments, side effects of medication, and importance of treatment adherence), stigma, recovery, and suicide prevention information and resources. [Multimedia Appendix 1](#) [9,11,49,50,53] presents the educational content appraisal questionnaire.

3. *Technical aspects and quality assurance of the app*, including usability, app credibility, in-app advertisements, privacy and security safeguards, and gamification.

Selection of Apps

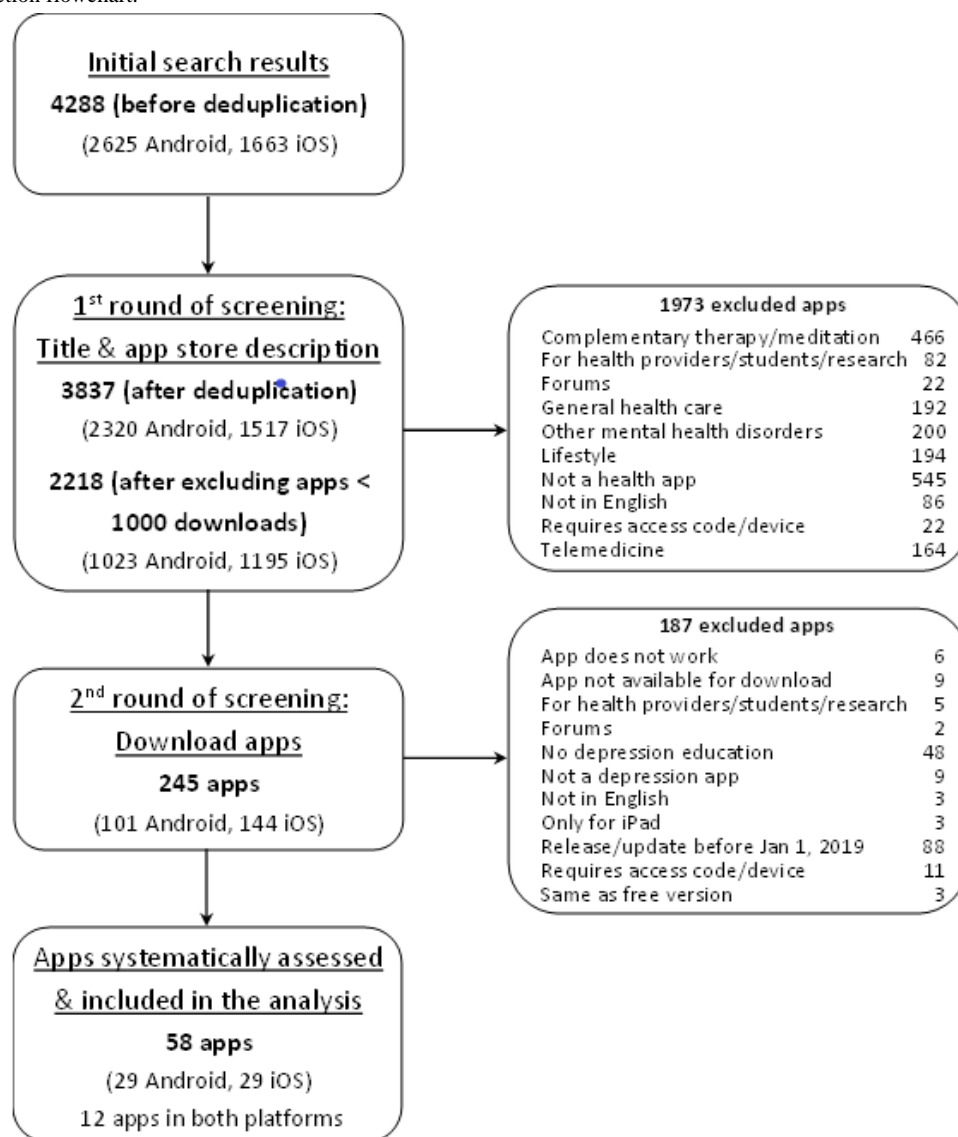
The Apple App Store and Google Play were systematically searched using 42matters, a proprietary app database [54], on July 8, 2020, using the terms *depression*, *depressive*, *depressed*, *mood disorder*, *sadness*, and *melancholia*. The search was limited to four app store categories: education, health and fitness, lifestyle, and medical. Eligible apps were required to conform to the criteria presented in [Textbox 1](#).

The app selection process is presented as a flowchart ([Figure 1](#)) [55]. Before screening for eligibility, we excluded all Android apps with <1000 downloads as reported in the 42matters search output. As iOS does not report the number of app downloads, the iOS versions of the excluded Android apps were also excluded on the assumption that the app would have a similar number of downloads in both app stores. The remaining apps underwent a 2-step selection process that consisted of (1) screening the app name and app store description from the 42 matters search output and (2) downloading and screening for eligibility all apps included in step 1. Working in parallel, 2 pairs of investigators (LM and Goh Jun Wei; LM and Matthew Teo Siu Yan) independently completed the app selection process. Disagreements were resolved through discussion.

Textbox 1. Inclusion and exclusion criteria for smartphone app selection.

Inclusion criteria
<ul style="list-style-type: none"> • Provides information about depression, including clinical presentation, diagnosis, and management (both pharmacological and nonpharmacological) • Targets depression or includes depression information within a general mental health app • Uploaded or updated from January 1, 2019, onward • Downloaded at least 1000 times • Available for free or requires payment to download, use, or expand functionalities (in-app purchases) and is available for download in the Apple App Store or Google Play • Available in English
Exclusion criteria
<ul style="list-style-type: none"> • Presents an overview of common physical and mental disorders, using a glossary format • Targets health care providers (eg, physicians, psychologists, and counselors) or the support network of a person with depression or consists exclusively of peer-support forums • Offers teleconsultation services with physicians, psychologists, counselors, or other health care providers • Consists of a stand-alone depression screening questionnaire, without education modules • Offers complementary medicine, meditation, or lifestyle improvement • Does not provide any depression-related information or includes non-health-related content (ie, music playlists, wallpapers, and so on) • Requires an access code to log in, or could not be used after 2 log in attempts because of technical problems, or was withdrawn from the app store at the time of access

Figure 1. App selection flowchart.



App Assessment and Data Analysis

Apps were assessed by 2 assessors working in parallel using an iPhone 7 (iOS version 14.0.1) and a Sony XPERIA XZs (Android version 8.0.0) smartphone. If apps were available on both platforms, as per the 42matters search output, we assessed both versions and counted each version as an individual app. We used descriptive statistics to analyze the data. We compiled and tabulated the results and reported them as a narrative synthesis.

Subgroup Analysis of Android Apps

We performed a subgroup analysis of Android apps to assess whether the number of app downloads was associated with the educational content or the quality or number of features offered by the apps. Apps were categorized into two groups: apps downloaded 1000-10,000 times and apps downloaded >10,000 times. The assessment included selected items from all 3 sections of the assessment. The data were tabulated and compared using a significance test for categorical variables: the chi-square test was used if each category contained >10 variables and a 2-tailed Fisher exact test was used if any of the categories

in the contingency table contained <10 variables. Statistical significance was set at $P < .05$. Statistical analyses were performed in RStudio (R version 4.0.3). iOS apps were not included in this analysis because the Apple App Store does not include this information.

Results

The app search retrieved 2218 results after removing duplicates and excluding apps with <1000 downloads, of which 245 were downloaded and 58 were included in the analysis. Figure 1 describes the app selection process.

General Characteristics of Apps

Table 1 presents a summary of the characteristics of the included apps. Of the 58 apps included in this analysis, 29 (50%) were Android apps and 29 (50%) were iOS apps; 21% (12/58) of the apps were available on both platforms, and 60% (35/58) of the apps belonged to the health and fitness app store category. Of the 29 Android apps, 3 (10%) had been downloaded >1 million times [56-58]. Of the 58 included apps, 15 (26%) offered only education and information modules, whereas 37 (64%) offered

education modules along with other mental health or depression management features. A simple interface that did not allow for user feedback or customization was offered by 41% (12/29) of

the Android apps and 31% (9/29) of the iOS apps. More than 10% of the apps offered education modules targeted to specific user groups.

Table 1. General characteristics of apps (N=58).

Feature	Android (n=29), n (%)	iOS (n=29), n (%)	Total (N=58), n (%)
App store category			
Education	4 (14)	0 (0)	4 (7)
Health and fitness	17 (59)	18 (62)	35 (60)
Lifestyle	1 (3)	1 (3)	2 (3)
Medical	7 (24)	10 (34)	17 (29)
App store rating, stars			
3.6 to 5	22 (76)	18 (62)	40 (69)
1 to 3.5	3 (10)	4 (14)	7 (12)
No ratings	4 (14)	7 (24)	11 (19)
App cost			
Free	18 (62)	15 (52)	33 (57)
Free + in-app purchases	11 (38)	14 (48)	25 (43)
Paid	0 (0)	0 (0)	0 (0)
Language			
English	26 (90)	23 (79)	49 (84)
English and other languages	3 (10)	6 (21)	9 (16)
Target user of the app			
No target user	25 (86)	26 (90)	51 (88)
Police officers	1 (3)	0 (0)	1 (2)
Veterans	1 (3)	1 (3)	2 (3)
Youth aged 12-18 years ^a	2 (7)	2 (7)	4 (7)
Scope of the app			
General mental health	21 (72)	18 (62)	39 (67)
Depression	8 (28)	11 (38)	19 (33)
Type of app			
Information and education	10 (34)	5 (17)	15 (26)
Disease management with education section	16 (55)	21 (72)	37 (64)
Education with disease management section	1 (3)	2 (7)	3 (5)
Multimedia education	2 (7)	1 (3)	3 (5)
Number of education topics			
<10	7 (24)	6 (21)	13 (22)
10-20	18 (62)	20 (69)	38 (66)
>20	4 (14)	3 (10)	7 (12)
Emergency contact information for users at risk of suicide	16 (55)	19 (66)	35 (60)
Peer-support communities	2 (7)	2 (7)	4 (7)
Non-evidence-based information	10 (34)	2 (7)	12 (21)

^aYouth was defined by the cutoff age provided by the apps.

Depression Education Modules

Overview

[Multimedia Appendix 2](#) presents a detailed description of the educational content in included apps. The apps offered a variety of information or educational topics, as summarized in [Table 2](#). Most of the apps (51/58, 88%) included up to 20 educational

topics. One in five apps provided non-evidence-based information, mostly in the form of personal opinions of the developers or columnists. Most of the apps informed users about the symptoms of depression and listed available treatments. Personal narratives on depression [51] were included only in approximately 20% of the apps.

Table 2. Depression education topics included in the apps (N=58).

Education topics included in the app	Android (n=29), n (%)	iOS (n=29), n (%)	Total (N=58), n (%)
General information on depression			
Personal narratives of depression	6 (21)	6 (21)	12 (21)
Depression is different from sadness	20 (69)	19 (66)	39 (67)
Demographic and epidemiological facts	15 (52)	18 (62)	33 (57)
Natural history of the disease	6 (21)	12 (41)	18 (31)
Lists symptoms of depression	26 (90)	26 (90)	52 (90)
Explains what recurrence and relapse are	9 (31)	5 (17)	14 (24)
Addresses stigma linked to depression	11 (38)	11 (38)	22 (38)
Mentions suicide risk linked to depression	23 (79)	20 (69)	43 (74)
Screening for depression			
Describes diagnostic criteria of depression	7 (24)	6 (21)	13 (22)
Provides reference to DSM-5 ^a or ICD-10 ^b	3 (10)	3 (10)	6 (10)
Administers a screening questionnaire			
PHQ-9 ^c	5 (45)	8 (50)	13 (48)
PHQ-9 + GAD-7 ^d	1 (9)	0 (0)	1 (4)
PHQ-9 + other validated questionnaires	1 (9)	2 (13)	3 (11)
Other validated questionnaires	1 (9)	2 (13)	3 (11)
Nonvalidated questionnaires	3 (27)	4 (25)	7 (26)
Explains the need for a confirmatory diagnosis after screening	7 (24)	3 (10)	10 (17)
Treatment of depression			
Importance of seeking help	23 (79)	21 (72)	44 (76)
Addresses phases and types of treatment (stepped or integrated treatment)	1 (3)	1 (3)	2 (3)
Advises to seek specialist treatment	22 (76)	21 (72)	43 (74)
Importance of involving support network	16 (55)	13(45)	29 (50)
Importance of complying with treatment	8 (28)	8 (28)	16 (28)
Lists available treatments	27 (93)	26 (90)	53 (91)
Treatments mentioned by the app			
Psychotherapy	3 (10)	4 (14)	7 (12)
Psychotherapy + pharmacotherapy	13 (45)	10 (34)	23 (40)
Psychotherapy + pharmacotherapy + others	7 (24)	7 (24)	14 (24)
Psychotherapy + others	0 (0)	2 (7)	2 (3)
Complementary medicine	3 (10)	1 (3)	4 (7)
Others	1 (3)	2 (7)	3 (5)
Importance of lifestyle changes	18 (62)	21 (72)	39 (67)
Effects of physical exercise	19 (66)	20 (69)	39 (67)
Offers emergency helpline phone numbers	16 (55)	19 (66)	35 (60)
Prognosis in patients receiving treatment	7 (24)	13 (45)	20 (34)
Addresses recovery and provides a hopeful outlook	3 (10)	3 (10)	6 (10)

^aDSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition.^bICD-10: International Classification of Diseases, Tenth Revision.^cPHQ-9: Patient Health Questionnaire, 9-item version.^dGAD-7: Generalized Anxiety Disorder, 7-item scale.

Symptoms and Natural History of Depression

Of the 58 included apps, 52 (90%) listed the symptoms of depression, whereas 39 (67%) stated the difference between depression and occasional low moods or sadness. Brief information on the epidemiology of depression was included in half of the assessed apps. Stigma associated with depression was mentioned by 38% (22/58) of the apps, whereas only 10% (6/58) included information on recovery.

Suicide risk associated with depression was mentioned by 74% (43/58) of the apps, generally as an item in a list of symptoms. Suicide prevention resources were offered by 60% (35/58) of the apps, including 55% (16/29) of the Android apps and 66% (19/29) of the iOS apps. The most common resource offered by the apps was information on crisis helplines (telephone numbers or website addresses).

Screening for Depression

Of the 58 included apps, 13 (22%) offered information on the diagnostic criteria for depression, 6 (10%) referenced psychiatric diagnostic manuals, and 10 (17%) explained the need to confirm diagnosis of depression after a positive screening test. Of the 58 apps, 27 (47%) apps administered a screening questionnaire for depression, commonly the Patient Health Questionnaire-9 [59]. Only 48% (13/27) of the apps responded to positive screening test scores by either offering self-management exercises for the user to engage with or suggesting consulting with a health care provider, commonly a physician, whereas the rest of the apps did not provide any feedback.

Treatment of Depression

Of the 58 included apps, 44 (76%) highlighted the importance of seeking help when affected by depression and consulting specialized providers if the symptoms were severe or persistent. Half of the Android apps and one-third of the iOS apps also emphasized the importance of involving members of the user's support network. Few apps included providers' contact information, and when such information was present, it was limited to contact information of the app development team; none of the apps included geolocated lists of health care providers by; for example, facilitating a search through navigation providers such as Google Maps.

More than 90% of the apps listed existing treatments for depression, without providing a description of what each category entailed. Psychotherapy was the most commonly mentioned treatment category (44/58, 76%), followed by medications (37/58, 64%), whereas 12% (7/58) of the apps included brain stimulation treatments as well. Nevertheless, few apps offered more comprehensive information on the different

treatment modalities, including explanations of what each treatment involves and relevant examples (eg, types of psychotherapy and antidepressant groups). Information on psychotherapy modalities was offered by 34% (10/29) of the Android apps and 52% (15/29) of the iOS apps, whereas few Android or iOS apps offered information on medication types (8/58, 14%), side effects (13/58, 22%), nonaddictive nature of frequently used antidepressants (5/58, 9%), or gradual initiation and discontinuation of therapy (10/58, 17%).

Benefits of lifestyle modifications, including regular exercise, was mentioned by 67% (39/58) of the assessed apps, whereas complementary medicine, including St John's wort, light therapy, and acupuncture, was mentioned by 7% (4/58) of the apps.

Technical Aspects and Quality Assurance of the App

All assessed apps worked as intended and were generally easy to use. Data entry from users was not requested by 31% (9/29) of the iOS apps and 41% (12/29) of the Android apps (Table 3).

Of the 58 included apps, 13 (22%) available on both platforms did not include a privacy policy (6/29, 21%, Android apps and 7/29, 24%, iOS apps). These apps offered passive information that did not require data entry by users, except for 3% (1/29) of the Android apps and 7% (2/29) of the iOS apps, which requested the user's email address to set up a password.

Of the 58 apps, 30 (52%) included references or the author's signature or shared links to reputable websites for the information provided. Half of the Android apps and a quarter of the iOS apps did not declare developers' affiliations, and 41% (24/58) of the apps included a disclaimer that the information provided did not replace a health care provider's advice.

Of the 58 apps, 10 (17%) contained advertisements. These apps were free to download and use, except for an Android app that offered in-app purchases. The advertisements included in 75% (6/8) of the Android apps filled the screen and disrupted the use of the app, whereas 25% (2/8) of the apps included banner advertisements that allowed users to continue using the app. All advertisements were generally unrelated to the app content, promoting other Google Play apps, video games, beauty products, education support centers, and health care services such as dental services. Of these 8 Android apps, 1 (13%) promoted psychological services. In contrast, 100% (2/2) of the iOS apps offered banner advertisements that allowed for continued use of the app. The advertisements were not related to the app content.

Table 3. Technical features and quality assurance of included apps (N=58).

Characteristics	Android (n=29), n (%)	iOS (n=29), n (%)	Total (N=58), n (%)
App credibility			
App content referenced or signed by the author	17 (59)	13 (45)	30 (52)
Includes disclaimer: information does not replace health care provider's advice	11 (38)	13 (45)	24 (41)
App development team included the following			
Government agency or academic institution or NGO ^a	4 (14)	8 (28)	12 (21)
Health care professional	11 (38)	14 (48)	25 (43)
Not declared	14 (48)	7 (24)	21 (36)
Data privacy			
Authentication required to access app	13 (45)	21 (72)	34 (59)
App includes a privacy policy			
Presented before account creation	14 (61)	10 (45)	24 (53)
Explains how data are collected	23 (100)	22 (100)	45 (100)
Shares information with third party providers	20 (87)	20 (91)	40 (89)
Contact details of data protection officer provided	4 (17)	9 (41)	13 (29)
App allows users to share data	6 (21)	7 (24)	13 (22)
In-app advertisements (for details, please see text in this section)	8 (28)	2 (7)	10 (17)

^aNGO: nongovernmental organization.

Subgroup Analysis of Android Apps

Our subgroup analysis assessing the association between the Android apps' popularity (as indexed by the number of app downloads) and the breadth and depth of the educational topics covered showed no statistically significant difference between apps downloaded <10,000 times and apps downloaded >10,000 times. Nonetheless, apps downloaded 1000-10,000 times

provided non-evidence-based information more often than apps downloaded >10,000 times (7/15, 47% vs 3/14, 21%, respectively); however, this difference was not statistically significant. Of the 7 apps providing non-evidence-based information and downloaded <10,000 times, 2 (29%) belonged to the medical category in the app store. Of these 2 apps, 1 (50%) claimed that adhering to the *Law of Attraction* improve depression outcomes. [Table 4](#) summarizes the analysis.

Table 4. Android apps subgroup analysis according to number of downloads (N=29).

Feature	>10,000 downloads (n=14), n (%)	1000-10,000 downloads (n=15), n (%)	<i>P</i> value ^a
App store category			.75
Education	1 (7)	3 (20)	
Health and fitness	9 (64)	8 (53)	
Lifestyle	0 (0)	1 (7)	
Medical	4 (29)	3 (20)	
App store rating (stars)			.09
3.6 to 5	13 (93)	9 (60)	
1 to 3.5	1 (7)	2 (13)	
No ratings	0 (0)	4 (27)	
App cost			.06
Free	6 (43)	12 (80)	
Free + in-app purchase	8 (57)	3 (20)	
Paid	0 (0)	0 (0)	
Type of app			.07
Information and education	2 (14)	8 (53)	
Disease management with education section	10 (71)	6 (40)	
Education disease management section	1 (7)	0 (0)	
Multimedia education	1 (7)	1 (7)	
Number of education topics			.66
<10	3 (21)	4 (27)	
10-20	8 (57)	10 (67)	
>20	3 (21)	1 (7)	
Peer-support communities	2 (14)	0 (0)	.22
Non-evidence-based information	3 (21)	7 (47)	.25
Depression education			
General information on depression			
Personal narratives of depression	4 (29)	2 (13)	.39
Depression is different from sadness	11 (79)	9 (60)	.43
Demographic and epidemiological facts	9 (64)	6 (40)	.35
Natural history of the disease	6 (43)	0 (0)	.006
Lists symptoms of depression	14 (100)	12 (80)	.22
Explains what recurrence and relapse are	4 (29)	5 (33)	.99
Addresses stigma linked to depression	7 (50)	4 (27)	.36
Mentions suicide risk linked to depression	12 (86)	11 (73)	.65
Screening of depression			
Describes diagnostic criteria of depression	4 (29)	3 (20)	.69
Administers a screening questionnaire	7 (50)	4 (27)	.36
Explains need for a confirmatory diagnosis after screening	4 (29)	3 (20)	.68
Treatment of depression			
Importance of seeking help	12 (86)	11 (73)	.65
Advises to seek specialist treatment	9 (64)	13 (87)	.21

Feature	>10,000 downloads (n=14), n (%)	1000-10,000 downloads (n=15), n (%)	<i>P value</i> ^a
Importance of involving support network	9 (64)	7 (47)	.56
Importance of complying with treatment	4 (29)	4 (27)	.99
Lists available treatments	11 (79)	12 (80)	.99
Importance of lifestyle changes	10 (71)	8 (53)	.53
Lists complementary medicine options	0 (0)	3 (20)	.22
Offers emergency helpline phone numbers	10 (71)	6 (40)	.18
Prognosis in patients receiving treatment	5 (36)	0 (0)	.02
Addresses recovery and provides a hopeful outlook	2 (14)	2 (13)	.99
App credibility			
App content referenced or signed by author	12 (86)	5 (33)	.012
Include disclaimer: information does not replace health care provider's advice	7 (50)	4 (27)	.36
App development team included the following			.11
Government agency or academic institution or NGO ^b	2 (14)	2 (13)	
Health care professional	8 (57)	3 (20)	
Not declared	4 (29)	10 (67)	
Data privacy			
Authentication required to access app	9 (64)	4 (27)	.10
App includes a privacy policy	12 (86)	11 (73)	.65
Presented before account creation	9 (64)	5 (33)	.20
Explains how data are collected	12 (86)	11 (73)	.65
Shares information with third party providers	10 (71)	10 (67)	.99
Contact details of data protection officer provided	4 (29)	0 (0)	.04
In-app advertisements	2 (14)	6 (40)	.21

^aStatistically significant values in italics.

^bNGO: nongovernmental organization.

Discussion

Principal Findings

To our knowledge, this is the first systematic assessment of the information on depression in existing mental health and depression apps. Our findings suggest that the information included in these apps is often limited and not aligned with evidence. Most of the assessed apps offered brief, factual information on symptoms and treatment options, with only a few apps addressing all aspects of disease presentation and management or including information on recovery or providing personal narratives of people living with depression. The inclusion of personal accounts of people living with, and recovering from, a mental health disorder has been associated with increased understanding of the mental health disorder and recovery process, validation of the personal experience, and reduction of stigma [28,60].

Patient education is one of the pillars of effective depression management resulting in better treatment compliance and improved outcomes [9,11], particularly if they are complemented

with personal narratives and positive messages emphasizing recovery [51]. Adequate health literacy is associated with tolerance and acceptance of people living with a mental disorder [16], increased help-seeking behavior [16], and adherence to treatment [61,62], and it is a prerequisite to engage in shared-decision models of care [63]. Yet, 25% of the potentially eligible apps, including some of the most popular mental health apps with >1 million downloads each, were excluded in the second round of screening because they did not provide any information on depression.

Access to mental health care is often inadequate, particularly in low- and middle-income countries [3] and among people of lower socioeconomic status as well as those living in remote areas in high-income countries [64,65].

Mental health apps may play an important role in improving access to mental health care, supported by an increasing expansion of mobile network coverage to remote areas worldwide [31]. However, our assessment shows that apps currently available in public app marketplaces do not offer a holistic, evidence-based self-management program to meet these

health needs. For example, less than half of the assessed apps include a screening test for depression and only a subset of these apps offer follow-up recommendations to users potentially living with depression. Of the 58 assessed apps, 12 (21%) offered non-evidence-based information, including developers' or columnists' personal views on depression, and suggested scientifically unproved treatments. Of these 12 apps, 2 (17%) were classified as *medical* in the app store. Lack of evidence-based information in health apps is a recurrent theme in most systematic assessments performed by our group [45,46,48] and others [42,66,67], revealing that current app development and publication processes overseen by app stores might not be suitable for health apps. We previously highlighted the lack of governance and quality assurance of the health app industry [48], which we believe constrains the further development of health apps as genuine tools to support access to mental health care. Therefore, it is imperative to develop a multifaceted approach involving the research community, commercial app developers, app store managers, and official regulatory bodies to define development and publication regulatory frameworks for health apps.

Our subgroup assessment of Android apps suggests an association between the number of downloads and the overall quality of the apps. In the analysis, apps downloaded >10,000 times were more trustworthy, included more evidence-based content, presented better compliance with data privacy and security, presented references to endorse claims, and incorporated health care providers into the app development team more often than apps downloaded <10,000 times. Excluding app credibility, the differences were not statistically significant, probably because of the small size of the sample.

In this study, we used an established systematic assessment approach, which is based on the rigorous systematic review and

which we have applied to various health domains [46-48] over the years. This methodology included the use of a commercial database to search for eligible apps, which provides a wider geographical scope for the search, and the development of exhaustive assessment criteria using renowned evidence-based clinical guidelines and our center's criteria for technical and quality assurance of apps.

There are limitations to our research as well. The search strategy included depression-related terms and omitted education-based terminology; thus, we may have omitted some relevant apps. However, the inclusion of such terms would have greatly increased the overall number of retrieved apps and would have also led to retrieval of many irrelevant apps because the database used for app retrieval does not allow the use of search strings; instead, it retrieves individual results for each search term. We restricted our search to apps for mental health, excluding apps offering medical or general health advice that may have presented summaries on depression presentation or management. Our assessment was limited to apps downloaded at least 1000 times, which may have omitted newly launched apps that have yet to reach the minimum required number of downloads. We also restricted our search to apps in English, potentially excluding relevant apps in languages other than English.

Conclusions

Information in mental health and depression apps is often brief and incomplete, with 1 in 5 apps providing non-evidence-based information. Given the unmet needs and stigma associated with depression, it is imperative that apps seize the opportunity to offer quality, evidence-based education and point users to relevant resources. A multistakeholder consensus on a more stringent development and publication process for mental health apps is essential.

Acknowledgments

The authors wish to thank Mr Matthew Teo Siu Yan and Mr Goh Jun Wei for their assistance in screening and assessing the apps. LM gratefully acknowledges Nanyang Technological University Research Scholarship (LKCmedicine) support for her PhD studentship which enabled this work. JL is supported by the Ministry of Health National Medical Research Council Clinician Scientist Award (MOH-000013). JC's post at Imperial College London is supported by the National Institute for Health Research Northwest London Applied Research Collaboration. JC is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise program.

Authors' Contributions

LM had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. LM, KG, and JC provided the concept and design of the study. LM acquired, analyzed, and interpreted the data; performed statistical analysis; and provided administrative, technical and material support. The manuscript was drafted by LM and ACS. All authors contributed for critical revision of the manuscript for important intellectual content. JC obtained the funding. KG and JC were responsible for supervision.

Conflicts of Interest

JL has received honoraria from Otsuka and Janssen. There are no other conflicts of interests.

Multimedia Appendix 1

Depression education topics assessment criteria.

[DOCX File, 37 KB - [jmir_v24i3e28942_app1.docx](#)]

Multimedia Appendix 2

Characteristics of included apps.

[\[DOCX File , 39 KB - jmir_v24i3e28942_app2.docx \]](#)**References**

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* 2018 Nov 10;392(10159):1789–1858 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(18\)32279-7](https://doi.org/10.1016/S0140-6736(18)32279-7)] [Medline: [30496104](https://pubmed.ncbi.nlm.nih.gov/30496104/)]
2. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. *Bull World Health Organ* 2004 Nov;82(11):858–866 [[FREE Full text](#)] [Medline: [15640922](https://pubmed.ncbi.nlm.nih.gov/15640922/)]
3. Luitel NP, Jordans MJ, Kohrt BA, Rathod SD, Komproe IH. Treatment gap and barriers for mental health care: a cross-sectional community survey in Nepal. *PLoS One* 2017;12(8):e0183223 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0183223](https://doi.org/10.1371/journal.pone.0183223)] [Medline: [28817734](https://pubmed.ncbi.nlm.nih.gov/28817734/)]
4. König H, König HH, Konnopka A. The excess costs of depression: a systematic review and meta-analysis. *Epidemiol Psychiatr Sci* 2019 Apr 05;29:e30 [[FREE Full text](#)] [doi: [10.1017/S2045796019000180](https://doi.org/10.1017/S2045796019000180)] [Medline: [30947759](https://pubmed.ncbi.nlm.nih.gov/30947759/)]
5. Jorm AF. Mental health literacy: empowering the community to take action for better mental health. *Am Psychol* 2012 Apr;67(3):231–243. [doi: [10.1037/a0025957](https://doi.org/10.1037/a0025957)] [Medline: [22040221](https://pubmed.ncbi.nlm.nih.gov/22040221/)]
6. Picco L, Abidin E, Pang S, Vaingankar JA, Jeyagurunathan A, Chong SA, et al. Association between recognition and help-seeking preferences and stigma towards people with mental illness. *Epidemiol Psychiatr Sci* 2018 Feb;27(1):84–93 [[FREE Full text](#)] [doi: [10.1017/S2045796016000998](https://doi.org/10.1017/S2045796016000998)] [Medline: [27927259](https://pubmed.ncbi.nlm.nih.gov/27927259/)]
7. Jorm AF, Christensen H, Griffiths KM. The public's ability to recognize mental disorders and their beliefs about treatment: changes in Australia over 8 years. *Aust N Z J Psychiatry* 2006 Jan;40(1):36–41. [doi: [10.1080/j.1440-1614.2006.01738.x](https://doi.org/10.1080/j.1440-1614.2006.01738.x)] [Medline: [16403035](https://pubmed.ncbi.nlm.nih.gov/16403035/)]
8. Jorm AF, Korten AE, Jacomb PA, Christensen H, Rodgers B, Pollitt P. "Mental health literacy": a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. *Med J Aust* 1997 Feb 17;166(4):182–186. [doi: [10.5694/j.1326-5377.1997.tb140071.x](https://doi.org/10.5694/j.1326-5377.1997.tb140071.x)] [Medline: [9066546](https://pubmed.ncbi.nlm.nih.gov/9066546/)]
9. Practice guideline for the treatment of patients with major depressive disorder (revision). American Psychiatric Association. *Am J Psychiatry* 2000 Apr;157(4 Suppl):1–45. [Medline: [10767867](https://pubmed.ncbi.nlm.nih.gov/10767867/)]
10. Depression The Treatment and Management of Depression in Adults (Updated Edition). Leicester (UK): British Psychological Society; 2010.
11. Malhi GS, Bassett D, Boyce P, Bryant R, Fitzgerald PB, Fritz K, et al. Royal Australian and New Zealand College of Psychiatrists clinical practice guidelines for mood disorders. *Aust N Z J Psychiatry* 2015 Dec;49(12):1087–1206. [doi: [10.1177/0004867415617657](https://doi.org/10.1177/0004867415617657)] [Medline: [26643054](https://pubmed.ncbi.nlm.nih.gov/26643054/)]
12. Menear M, Dugas M, Careau E, Chouinard M, Dogba MJ, Gagnon M, et al. Strategies for engaging patients and families in collaborative care programs for depression and anxiety disorders: a systematic review. *J Affect Disord* 2020 Feb 15;263:528–539 [[FREE Full text](#)] [doi: [10.1016/j.jad.2019.11.008](https://doi.org/10.1016/j.jad.2019.11.008)] [Medline: [31744737](https://pubmed.ncbi.nlm.nih.gov/31744737/)]
13. Slade M. Implementing shared decision making in routine mental health care. *World Psychiatry* 2017 Jun;16(2):146–153 [[FREE Full text](#)] [doi: [10.1002/wps.20412](https://doi.org/10.1002/wps.20412)] [Medline: [28498575](https://pubmed.ncbi.nlm.nih.gov/28498575/)]
14. Donker T, Griffiths KM, Cuijpers P, Christensen H. Psychoeducation for depression, anxiety and psychological distress: a meta-analysis. *BMC Med* 2009 Dec 16;7:79 [[FREE Full text](#)] [doi: [10.1186/1741-7015-7-79](https://doi.org/10.1186/1741-7015-7-79)] [Medline: [20015347](https://pubmed.ncbi.nlm.nih.gov/20015347/)]
15. Griffiths KM, Christensen H. Internet-based mental health programs: a powerful tool in the rural medical kit. *Aust J Rural Health* 2007 Apr;15(2):81–87. [doi: [10.1111/j.1440-1584.2007.00859.x](https://doi.org/10.1111/j.1440-1584.2007.00859.x)] [Medline: [17441815](https://pubmed.ncbi.nlm.nih.gov/17441815/)]
16. Rüschi N, Evans-Lacko SE, Henderson C, Flach C, Thornicroft G. Knowledge and attitudes as predictors of intentions to seek help for and disclose a mental illness. *Psychiatr Serv* 2011 Jun;62(6):675–678. [doi: [10.1176/ps.62.6.pss6206_0675](https://doi.org/10.1176/ps.62.6.pss6206_0675)] [Medline: [21632739](https://pubmed.ncbi.nlm.nih.gov/21632739/)]
17. Wang J, Lai D. The relationship between mental health literacy, personal contacts and personal stigma against depression. *J Affect Disord* 2008 Sep;110(1-2):191–196. [doi: [10.1016/j.jad.2008.01.005](https://doi.org/10.1016/j.jad.2008.01.005)] [Medline: [18261806](https://pubmed.ncbi.nlm.nih.gov/18261806/)]
18. Tonsing KN. A review of mental health literacy in Singapore. *Soc Work Health Care* 2018 Jan;57(1):27–47. [doi: [10.1080/00981389.2017.1383335](https://doi.org/10.1080/00981389.2017.1383335)] [Medline: [28976296](https://pubmed.ncbi.nlm.nih.gov/28976296/)]
19. Bonabi H, Müller M, Ajdacic-Gross V, Eisele J, Rodgers S, Seifritz E, et al. Mental health literacy, attitudes to help seeking, and perceived need as predictors of mental health service use: a longitudinal study. *J Nerv Ment Dis* 2016 Apr;204(4):321–324. [doi: [10.1097/NMD.0000000000000488](https://doi.org/10.1097/NMD.0000000000000488)] [Medline: [27015396](https://pubmed.ncbi.nlm.nih.gov/27015396/)]
20. Zhuang XY, Wong DF, Cheng C, Pan S. Mental health literacy, stigma and perception of causation of mental illness among Chinese people in Taiwan. *Int J Soc Psychiatry* 2017 Sep;63(6):498–507. [doi: [10.1177/0020764017719303](https://doi.org/10.1177/0020764017719303)] [Medline: [28805154](https://pubmed.ncbi.nlm.nih.gov/28805154/)]
21. Cheng H, Wang C, McDermott RC, Kridel M, Rislin JL. Self-stigma, mental health literacy, and attitudes toward seeking psychological help. *J Counseling Develop* 2018 Jan 04;96(1):64–74. [doi: [10.1002/jcad.12178](https://doi.org/10.1002/jcad.12178)]

22. Henderson C, Noblett J, Parke H, Clement S, Caffrey A, Gale-Grant O, et al. Mental health-related stigma in health care and mental health-care settings. *Lancet Psychiatry* 2014 Nov;1(6):467-482. [doi: [10.1016/S2215-0366\(14\)00023-6](https://doi.org/10.1016/S2215-0366(14)00023-6)] [Medline: [26361202](https://pubmed.ncbi.nlm.nih.gov/26361202/)]
23. Knaak S, Mantler E, Szeto A. Mental illness-related stigma in healthcare: barriers to access and care and evidence-based solutions. *Health Manage Forum* 2017 Mar;30(2):111-116 [FREE Full text] [doi: [10.1177/0840470416679413](https://doi.org/10.1177/0840470416679413)] [Medline: [28929889](https://pubmed.ncbi.nlm.nih.gov/28929889/)]
24. Corrigan PW, Watson AC. Understanding the impact of stigma on people with mental illness. *World Psychiatry* 2002 Feb;1(1):16-20 [FREE Full text] [Medline: [16946807](https://pubmed.ncbi.nlm.nih.gov/16946807/)]
25. Clement S, Schauman O, Graham T, Maggioni F, Evans-Lacko S, Bezborodovs N, et al. What is the impact of mental health-related stigma on help-seeking? A systematic review of quantitative and qualitative studies. *Psychol Med* 2015 Jan;45(1):11-27. [doi: [10.1017/S0033291714000129](https://doi.org/10.1017/S0033291714000129)] [Medline: [24569086](https://pubmed.ncbi.nlm.nih.gov/24569086/)]
26. Jorm AF, Christensen H, Griffiths KM. The impact of beyondblue: the national depression initiative on the Australian public's recognition of depression and beliefs about treatments. *Aust N Z J Psychiatry* 2005 Apr;39(4):248-254. [doi: [10.1080/j.1440-1614.2005.01561.x](https://doi.org/10.1080/j.1440-1614.2005.01561.x)] [Medline: [15777361](https://pubmed.ncbi.nlm.nih.gov/15777361/)]
27. Depression: let's talk campaign essentials. World Health Organization. URL: <https://www.who.int/campaigns/world-health-day/2017/toolkit.pdf?ua=1> [accessed 2021-01-21]
28. Henderson C, Evans-Lacko S, Thornicroft G. Mental illness stigma, help seeking, and public health programs. *Am J Public Health* 2013 May;103(5):777-780. [doi: [10.2105/AJPH.2012.301056](https://doi.org/10.2105/AJPH.2012.301056)] [Medline: [23488489](https://pubmed.ncbi.nlm.nih.gov/23488489/)]
29. Evans-Lacko S, Corker E, Williams P, Henderson C, Thornicroft G. Effect of the Time to Change anti-stigma campaign on trends in mental-illness-related public stigma among the English population in 2003-13: an analysis of survey data. *Lancet Psychiatry* 2014 Jul;1(2):121-128 [FREE Full text] [doi: [10.1016/S2215-0366\(14\)70243-3](https://doi.org/10.1016/S2215-0366(14)70243-3)] [Medline: [26360575](https://pubmed.ncbi.nlm.nih.gov/26360575/)]
30. Montagni I, Donisi V, Tedeschi F, Parizot I, Motrico E, Horgan A. Internet use for mental health information and support among European university students: the e-MentH project. *Digit Health* 2016;2:2055207616653845 [FREE Full text] [doi: [10.1177/2055207616653845](https://doi.org/10.1177/2055207616653845)] [Medline: [29942558](https://pubmed.ncbi.nlm.nih.gov/29942558/)]
31. Measuring digital development: facts and figures. International Telecommunications Union. URL: <https://www.itu.int/en/ITU-D/Statistics/Documents/facts/FactsFigures2020.pdf> [accessed 2021-02-09]
32. Timmers T, Janssen L, Kool RB, Kremer JA. Educating patients by providing timely information using smartphone and tablet apps: systematic review. *J Med Internet Res* 2020 Apr 13;22(4):e17342 [FREE Full text] [doi: [10.2196/17342](https://doi.org/10.2196/17342)] [Medline: [32281936](https://pubmed.ncbi.nlm.nih.gov/32281936/)]
33. Number of apps available in leading app stores as of 1st quarter 2021. Statista. URL: <https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/> [accessed 2022-02-23]
34. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health: transparency and trust. *JAMA Psychiatry* 2017 May 01;74(5):437-438. [doi: [10.1001/jamapsychiatry.2017.0262](https://doi.org/10.1001/jamapsychiatry.2017.0262)] [Medline: [28384700](https://pubmed.ncbi.nlm.nih.gov/28384700/)]
35. Van Ameringen M, Turna J, Khalesi Z, Pullia K, Patterson B. There is an app for that! The current state of mobile applications (apps) for DSM-5 obsessive-compulsive disorder, posttraumatic stress disorder, anxiety and mood disorders. *Depress Anxiety* 2017 Jun;34(6):526-539. [doi: [10.1002/da.22657](https://doi.org/10.1002/da.22657)] [Medline: [28569409](https://pubmed.ncbi.nlm.nih.gov/28569409/)]
36. Shang J, Wei S, Jin J, Zhang P. Mental health apps in China: analysis and quality assessment. *JMIR Mhealth Uhealth* 2019 Nov 07;7(11):e13236 [FREE Full text] [doi: [10.2196/13236](https://doi.org/10.2196/13236)] [Medline: [31697245](https://pubmed.ncbi.nlm.nih.gov/31697245/)]
37. Leigh S, Flatt S. App-based psychological interventions: friend or foe? *Evid Based Ment Health* 2015 Nov;18(4):97-99. [doi: [10.1136/eb-2015-102203](https://doi.org/10.1136/eb-2015-102203)] [Medline: [26459466](https://pubmed.ncbi.nlm.nih.gov/26459466/)]
38. Shen N, Levitan MJ, Johnson A, Bender JL, Hamilton-Page M, Jadad AA, et al. Finding a depression app: a review and content analysis of the depression app marketplace. *JMIR Mhealth Uhealth* 2015 Feb 16;3(1):e16 [FREE Full text] [doi: [10.2196/mhealth.3713](https://doi.org/10.2196/mhealth.3713)] [Medline: [25689790](https://pubmed.ncbi.nlm.nih.gov/25689790/)]
39. Baumel A, Torous J, Edan S, Kane JM. There is a non-evidence-based app for that: a systematic review and mixed methods analysis of depression- and anxiety-related apps that incorporate unrecognized techniques. *J Affect Disord* 2020 Aug 01;273:410-421. [doi: [10.1016/j.jad.2020.05.011](https://doi.org/10.1016/j.jad.2020.05.011)] [Medline: [32560936](https://pubmed.ncbi.nlm.nih.gov/32560936/)]
40. Qu C, Sas C, Daudén Roquet C, Doherty G. Functionality of top-rated mobile apps for depression: systematic search and evaluation. *JMIR Ment Health* 2020 Jan 24;7(1):e15321 [FREE Full text] [doi: [10.2196/15321](https://doi.org/10.2196/15321)] [Medline: [32012079](https://pubmed.ncbi.nlm.nih.gov/32012079/)]
41. Wasil AR, Venturo-Conerly KE, Shingleton RM, Weisz JR. A review of popular smartphone apps for depression and anxiety: assessing the inclusion of evidence-based content. *Behav Res Ther* 2019 Dec;123:103498. [doi: [10.1016/j.brat.2019.103498](https://doi.org/10.1016/j.brat.2019.103498)] [Medline: [31707224](https://pubmed.ncbi.nlm.nih.gov/31707224/)]
42. Bubolz S, Mayer G, Gronewold N, Hilbel T, Schultz JH. Adherence to established treatment guidelines among unguided digital interventions for depression: quality evaluation of 28 web-based programs and mobile apps. *J Med Internet Res* 2020 Jul 13;22(7):e16136 [FREE Full text] [doi: [10.2196/16136](https://doi.org/10.2196/16136)] [Medline: [32673221](https://pubmed.ncbi.nlm.nih.gov/32673221/)]
43. Tinschert P, Jakob R, Barata F, Kramer J, Kowatsch T. The potential of mobile apps for improving asthma self-management: a review of publicly available and well-adopted asthma apps. *JMIR Mhealth Uhealth* 2017 Aug 02;5(8):e113 [FREE Full text] [doi: [10.2196/mhealth.7177](https://doi.org/10.2196/mhealth.7177)] [Medline: [28768606](https://pubmed.ncbi.nlm.nih.gov/28768606/)]

44. Leigh S, Ouyang J, Mimmagh C. Effective? Engaging? Secure? Applying the ORCHA-24 framework to evaluate apps for chronic insomnia disorder. *Evid Based Ment Health* 2017 Nov;20(4):e20. [doi: [10.1136/eb-2017-102751](https://doi.org/10.1136/eb-2017-102751)] [Medline: [28947676](https://pubmed.ncbi.nlm.nih.gov/28947676/)]
45. 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](https://doi.org/10.1186/s12916-015-0314-7)] [Medline: [25943590](https://pubmed.ncbi.nlm.nih.gov/25943590/)]
46. Lum E, Jimenez G, Huang Z, Thai L, Semwal M, Boehm BO, et al. Decision support and alerts of apps for self-management of blood glucose for type 2 diabetes. *JAMA* 2019 Apr 16;321(15):1530-1532 [FREE Full text] [doi: [10.1001/jama.2019.1644](https://doi.org/10.1001/jama.2019.1644)] [Medline: [30990543](https://pubmed.ncbi.nlm.nih.gov/30990543/)]
47. van Galen LS, Xu X, Koh MJ, Thng S, Car J. Eczema apps conformance with clinical guidelines: a systematic assessment of functions, tools and content. *Br J Dermatol* 2020 Feb;182(2):444-453. [doi: [10.1111/bjd.18152](https://doi.org/10.1111/bjd.18152)] [Medline: [31179535](https://pubmed.ncbi.nlm.nih.gov/31179535/)]
48. Martinengo L, Van Galen L, Lum E, Kowalski M, Subramaniam M, Car J. Suicide prevention and depression apps' suicide risk assessment and management: a systematic assessment of adherence to clinical guidelines. *BMC Med* 2019 Dec 19;17(1):231 [FREE Full text] [doi: [10.1186/s12916-019-1461-z](https://doi.org/10.1186/s12916-019-1461-z)] [Medline: [31852455](https://pubmed.ncbi.nlm.nih.gov/31852455/)]
49. Depression. Ministry of Health Singapore. URL: https://www.moh.gov.sg/hpp/doctors/guidelines/GuidelineDetails/cpgmed_depression [accessed 2022-02-23]
50. Depression in adults: recognition and management. National Institute for Health and Care Excellence. 2009. URL: <http://www.nice.org.uk/guidance/cg90> [accessed 2022-02-23]
51. Clement S, Jarrett M, Henderson C, Thornicroft G. Messages to use in population-level campaigns to reduce mental health-related stigma: consensus development study. *Epidemiol Psychiatr Soc* 2010;19(1):72-79. [doi: [10.1017/s1121189x00001627](https://doi.org/10.1017/s1121189x00001627)] [Medline: [20486426](https://pubmed.ncbi.nlm.nih.gov/20486426/)]
52. Gronholm PC, Henderson C, Deb T, Thornicroft G. Interventions to reduce discrimination and stigma: the state of the art. *Soc Psychiatry Psychiatr Epidemiol* 2017 Mar;52(3):249-258 [FREE Full text] [doi: [10.1007/s00127-017-1341-9](https://doi.org/10.1007/s00127-017-1341-9)] [Medline: [28144713](https://pubmed.ncbi.nlm.nih.gov/28144713/)]
53. Thornicroft G, Mehta N, Clement S, Evans-Lacko S, Doherty M, Rose D, et al. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. *Lancet* 2016;387(10023):1123-1132. [doi: [10.1016/S0140-6736\(15\)00298-6](https://doi.org/10.1016/S0140-6736(15)00298-6)] [Medline: [26410341](https://pubmed.ncbi.nlm.nih.gov/26410341/)]
54. All the app intelligence data you need. 42matters. URL: <https://42matters.com/> [accessed 2022-02-23]
55. 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](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
56. Google Play. URL: https://play.google.com/store/apps/details?id=com.sevencupsoftea.app&hl=en_SG [accessed 2022-02-23]
57. MindDoc companion. Google Play. URL: https://play.google.com/store/apps/details?id=de.moodpath.android&hl=en_SG [accessed 2022-02-23]
58. Youper therapy. Google Play. URL: https://play.google.com/store/apps/details?id=br.com.youper&hl=en_SG [accessed 2022-02-23]
59. 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](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
60. Rennick-Egglestone S, Morgan K, Llewellyn-Beardsley J, Ramsay A, McGranahan R, Gillard S, et al. Mental health recovery narratives and their impact on recipients: systematic review and narrative synthesis. *Can J Psychiatry* 2019 Oct;64(10):669-679 [FREE Full text] [doi: [10.1177/0706743719846108](https://doi.org/10.1177/0706743719846108)] [Medline: [31046432](https://pubmed.ncbi.nlm.nih.gov/31046432/)]
61. Miller TA. Health literacy and adherence to medical treatment in chronic and acute illness: a meta-analysis. *Patient Educ Couns* 2016 Jul;99(7):1079-1086 [FREE Full text] [doi: [10.1016/j.pec.2016.01.020](https://doi.org/10.1016/j.pec.2016.01.020)] [Medline: [26899632](https://pubmed.ncbi.nlm.nih.gov/26899632/)]
62. Bauer AM, Schillinger D, Parker MM, Katon W, Adler N, Adams AS, et al. Health literacy and antidepressant medication adherence among adults with diabetes: the diabetes study of Northern California (DISTANCE). *J Gen Intern Med* 2013 Sep;28(9):1181-1187 [FREE Full text] [doi: [10.1007/s11606-013-2402-8](https://doi.org/10.1007/s11606-013-2402-8)] [Medline: [23512335](https://pubmed.ncbi.nlm.nih.gov/23512335/)]
63. Muscat DM, Shepherd HL, Nutbeam D, Trevena L, McCaffery KJ. Health literacy and shared decision-making: exploring the relationship to enable meaningful patient engagement in healthcare. *J Gen Intern Med* 2021 Feb;36(2):521-524 [FREE Full text] [doi: [10.1007/s11606-020-05912-0](https://doi.org/10.1007/s11606-020-05912-0)] [Medline: [32472490](https://pubmed.ncbi.nlm.nih.gov/32472490/)]
64. Fundamental Facts about mental health 2016. Mental Health Foundation. 2016. URL: <https://www.mentalhealth.org.uk/publications/fundamental-facts-about-mental-health-2016> [accessed 2020-06-17]
65. Kirby JB, Zuvekas SH, Borsky AE, Ngo-Metzger Q. Rural residents with mental health needs have fewer care visits than urban counterparts. *Health Aff (Millwood)* 2019 Dec;38(12):2057-2060. [doi: [10.1377/hlthaff.2019.00369](https://doi.org/10.1377/hlthaff.2019.00369)] [Medline: [31794321](https://pubmed.ncbi.nlm.nih.gov/31794321/)]
66. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. *J Med Internet Res* 2015 Aug 17;17(8):e198 [FREE Full text] [doi: [10.2196/jmir.4581](https://doi.org/10.2196/jmir.4581)] [Medline: [26283290](https://pubmed.ncbi.nlm.nih.gov/26283290/)]
67. Wisniewski H, Liu G, Henson P, Vaidyam A, Hajratalli NK, Onnela J, et al. Understanding the quality, effectiveness and attributes of top-rated smartphone health apps. *Evid Based Ment Health* 2019 Feb;22(1):4-9 [FREE Full text] [doi: [10.1136/ebmental-2018-300069](https://doi.org/10.1136/ebmental-2018-300069)] [Medline: [30635262](https://pubmed.ncbi.nlm.nih.gov/30635262/)]

Edited by R Kukafka; submitted 22.03.21; peer-reviewed by K Schroeder, KL Mauco; comments to author 26.04.21; revised version received 08.05.21; accepted 16.12.21; published 09.03.22.

Please cite as:

Martinengo L, Stona AC, Tudor Car L, Lee J, Griva K, Car J

Education on Depression in Mental Health Apps: Systematic Assessment of Characteristics and Adherence to Evidence-Based Guidelines

J Med Internet Res 2022;24(3):e28942

URL: <https://www.jmir.org/2022/3/e28942>

doi: [10.2196/28942](https://doi.org/10.2196/28942)

PMID: [35262489](https://pubmed.ncbi.nlm.nih.gov/35262489/)

©Laura Martinengo, Anne-Claire Stona, Lorainne Tudor Car, Jimmy Lee, Konstadina Griva, Josip Car. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 09.03.2022. 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.

Original Paper

Effect of Sleep Disturbance Symptoms on Treatment Outcome in Blended Cognitive Behavioral Therapy for Depression (E-COMPARED Study): Secondary Analysis

Esben Skov Jensen^{1,2}, MSc; Nicolai Ladegaard³, PhD; Angelina Isabella Mellentin^{1,2,4,5}, PhD; David Daniel Ebert⁶, PhD; Ingrid Titzler⁷, PhD; Ricardo Araya⁸, PhD; Arlinda Cerga Pashoja^{9,10}, PhD; Jean-Baptiste Hazo^{11,12}, PhD; Jérôme Holtzmann^{13,14}, PhD; Roman Cieslak^{15,16}, PhD; Ewelina Smoktunowicz¹⁵, PhD; Rosa Baños^{17,18}, PhD; Rocio Herrero^{17,18}, PhD; Azucena García-Palacios^{18,19}, PhD; Cristina Botella^{18,19}, PhD; Thomas Berger²⁰, PhD; Tobias Krieger²⁰, PhD; Trine Theresa Holmberg^{1,2}, MSc; Naira Topooco^{21,22}, PhD; Gerhard Andersson^{21,23,24}, PhD; Annemieke van Straten^{25,26}, PhD; Lise Kemmeren²⁵, PhD; Annet Kleiboer²⁵, PhD; Heleen Riper^{1,25}, PhD; Kim Mathiasen^{1,2,3}, PhD

¹Centre for Telepsychiatry, Mental Health Services of Southern Denmark, Odense, Denmark

²Department of Clinical Medicine, Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark

³Department of Affective Disorders, Aarhus University Hospital, Aarhus, Denmark

⁴Unit for Psychiatric Research, Department of Clinical Research, University of Southern Denmark, Odense, Denmark

⁵Brain Research-Inter-Disciplinary Guided Excellence (BRIDGE), Department of Clinical Research, University of Southern Denmark, Odense, Denmark

⁶Department for Sport and Health Sciences, Chair for Psychology & Digital Mental Health Care, Technical University Munich, Munich, Germany

⁷Department of Clinical Psychology and Psychotherapy, Friedrich-Alexander University of Erlangen-Nürnberg, Erlangen, Germany

⁸Centre for Global Mental Health, King's College London, London, United Kingdom

⁹Department of Population Health, London School of Hygiene and Tropical Medicine, London, United Kingdom

¹⁰Global Public Health, Public Health England, London, United Kingdom

¹¹URC Eco Ile-de-France (AP-HP), Hotel Dieu, Paris, France

¹²Assistance Publique-Hôpitaux de Paris, Health Economics Research Unit, University of Paris, Paris, France

¹³Fondation FondaMental, Créteil, France

¹⁴Service de Psychiatrie de l'adulte, Centre Expert Dépression Résistante FondaMental, Grenoble University Hospital, Grenoble, France

¹⁵Faculty of Psychology, SWPS University of Social Sciences and Humanities, Warsaw, Poland

¹⁶Lyda Hill Institute for Human Resilience, University of Colorado, Colorado Springs, CO, United States

¹⁷Instituto Polibienestar, University of Valencia, Valencia, Spain

¹⁸CIBER Fisiopatología Obesidad y Nutrición (CIBEROBN), Madrid, Spain

¹⁹Department of Basic Psychology, Clinic and Psychobiology, University Jaume I, Castellón, Spain

²⁰Department of Clinical Psychology and Psychotherapy, University of Bern, Bern, Switzerland

²¹Department of Behavioural Sciences and Learning, Linköping University, Linköping, Sweden

²²Center for m2Health, Palo Alto University, Palo Alto, CA, United States

²³Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

²⁴Department of Clinical Neuroscience, Karolinska Institute, Stockholm, Sweden

²⁵Department of Clinical, Neuro, and Developmental Psychology, Vrije Universiteit, Amsterdam, Netherlands

²⁶Amsterdam Public Health Research Institute, Amsterdam, Netherlands

Corresponding Author:

Kim Mathiasen, PhD

Centre for Telepsychiatry

Mental Health Services of Southern Denmark

Odense

Denmark

Phone: 1 61677747

Email: kmathiasen@health.sdu.dk

Abstract

Background: Sleep disturbance symptoms are common in major depressive disorder (MDD) and have been found to hamper the treatment effect of conventional face-to-face psychological treatments such as cognitive behavioral therapy. To increase the dissemination of evidence-based treatment, blended cognitive behavioral therapy (bCBT) consisting of web-based and face-to-face treatment is on the rise for patients with MDD. To date, no study has examined whether sleep disturbance symptoms have an impact on bCBT treatment outcomes and whether it affects bCBT and treatment-as-usual (TAU) equally.

Objective: The objectives of this study are to investigate whether baseline sleep disturbance symptoms have an impact on treatment outcomes independent of treatment modality and whether sleep disturbance symptoms impact bCBT and TAU in routine care equally.

Methods: The study was based on data from the E-COMPARED (European Comparative Effectiveness Research on Blended Depression Treatment Versus Treatment-as-Usual) study, a 2-arm, multisite, parallel randomized controlled, noninferiority trial. A total of 943 outpatients with MDD were randomized to either bCBT (476/943, 50.5%) or TAU consisting of routine clinical MDD treatment (467/943, 49.5%). The primary outcome of this study was the change in depression symptom severity at the 12-month follow-up. The secondary outcomes were the change in depression symptom severity at the 3- and 6-month follow-up and MDD diagnoses at the 12-month follow-up, assessed using the Patient Health Questionnaire-9 and Mini-International Neuropsychiatric Interview, respectively. Mixed effects models were used to examine the association of sleep disturbance symptoms with treatment outcome and treatment modality over time.

Results: Of the 943 patients recruited for the study, 558 (59.2%) completed the 12-month follow-up assessment. In the total sample, baseline sleep disturbance symptoms did not significantly affect change in depressive symptom severity at the 12-month follow-up ($\beta=.16$, 95% CI -0.04 to 0.36). However, baseline sleep disturbance symptoms were negatively associated with treatment outcome for bCBT ($\beta=.49$, 95% CI 0.22 - 0.76) but not for TAU ($\beta=-.23$, 95% CI -0.50 to 0.05) at the 12-month follow-up, even when adjusting for baseline depression symptom severity. The same result was seen for the effect of sleep disturbance symptoms on the presence of depression measured with Mini-International Neuropsychiatric Interview at the 12-month follow-up. However, for both treatment formats, baseline sleep disturbance symptoms were not associated with depression symptom severity at either the 3- ($\beta=.06$, 95% CI -0.11 to 0.23) or 6-month ($\beta=.09$, 95% CI -0.10 to 0.28) follow-up.

Conclusions: Baseline sleep disturbance symptoms may have a negative impact on long-term treatment outcomes in bCBT for MDD. This effect was not observed for TAU. These findings suggest that special attention to sleep disturbance symptoms might be warranted when MDD is treated with bCBT. Future studies should investigate the effect of implementing modules specifically targeting sleep disturbance symptoms in bCBT for MDD to improve long-term prognosis.

(*J Med Internet Res* 2022;24(3):e30231) doi:[10.2196/30231](https://doi.org/10.2196/30231)

KEYWORDS

blended care; bCBT; cognitive behavioral therapy; digital intervention; major depressive disorder; sleep disturbance; sleep disorder; mental health; digital health; mobile phone

Introduction

Major Depressive Disorder

Major depressive disorder (MDD) is among the most prevalent and debilitating psychiatric disorders [1-3], and it has been estimated to be the third leading cause of disability worldwide [3]. The MDD burden is particularly prominent in Western societies and poses an immense burden on society [1,3,4]. Although MDD, in many cases, can be successfully treated with evidence-based psychological and pharmacological therapies [5], some patients do not respond to treatment or do not receive adequate treatment [6]. To alleviate the burden of the disorder, we need to understand the mechanisms that may impact the lack of treatment response and develop strategies for broader dissemination of evidence-based psychological treatments [5,7].

Depression and Sleep Disturbance

MDD is a heterogeneous condition with a variety of presentations and a broad constellation of associated symptoms. Sleep disturbances are core symptoms of MDD, covering mostly insomnia (difficulties in falling asleep, wakefulness after sleep onset, and/or waking up early and not being able to get back to sleep) and, to a lesser extent, hypersomnia (excessive daytime

sleepiness and/or excessive total sleep time) [8]. Although insomnia and hypersomnia are common symptoms of MDD, they may also be diagnosed as independent psychiatric disorders [9]. Insomnia and mood disorders seem to have a bidirectional relationship, with insomnia increasing the risk for MDD, although MDD also increases the risk for insomnia symptoms [8]. This complex relationship seems to be founded on both common predisposing biological factors and cognitive and behavioral elements perpetuating both disorders, such as attentional biases and conditioning of arousal and negative affect in the bedroom [10]. Several studies have examined the impact of insomnia on the effect of MDD treatment and have consistently shown that insomnia affects the response to evidence-based treatment for depression, such as cognitive behavioral therapy (CBT) [11-14]. A total of 2 studies [12,14] reported that insomnia symptoms may double the risk for nonremission following depression treatment. In another study, patients exhibiting a limited response to treatment were more likely to present with recurring insomnia and depressive episodes, whereas optimal responders were consistently lower on insomnia measures [11]. In addition, a randomized controlled trial on older patients with MDD found that patients who presented with persistent insomnia symptoms, that is, reporting

symptoms at baseline and the 3-month follow-up, were less likely to reach remission at the 6- and 12-month time points [13]. Furthermore, insomnia symptoms are among the most frequent residual symptoms following depression treatment [15-17]. Hypersomnia has not received the same attention as comorbid insomnia in MDD, perhaps because it is less common [18]. However, studies have found that persons with MDD who experience both hypersomnia and insomnia are more severely depressed compared with individuals with depression with only insomnia or no sleep disturbance symptoms [18,19].

Digital Interventions

Digital interventions are often recommended as a tool to increase the dissemination of evidence-based psychological treatments [5,7]. Meta-analyses have shown that internet-based, guided self-help CBT (iCBT) is effective for treating depressive symptoms [20,21], and it may allow patients to circumvent some of the financial and structural barriers to psychological treatment [22]. However, iCBT is also affected by sleep disturbance symptoms. A cohort study on iCBT for depression in routine care showed that more sleep problems predicted higher depression by the end of treatment [23]. Furthermore, a randomized controlled trial on iCBT for chronic stress showed that the treatment effect was mediated by the reduction in insomnia symptoms [24]. Interestingly, some studies have shown that iCBT for insomnia also has significant effects on depressive symptoms [25,26], and in some cases, studies have also included insomnia management in the iCBT treatment manual for depression [27]. However, the added effect of doing so is yet to be documented.

As the self-help format in iCBT has a more rigid treatment structure, it might be more challenging to adapt the treatment to the individual's symptom presentation. Standard CBT protocols often do not target sleep disturbance symptoms during treatment. However, within the face-to-face consultation, the therapist may have more flexibility to address other issues experienced by the patient. Perhaps even more so in routine care treatments, where therapists are not required to adhere strictly to a protocol as in a research project. The therapist could also favor another treatment approach that emphasizes sleep disturbance symptoms more than CBT. Digital interventions might lack this flexibility, when some or all of the treatment is delivered in a standardized format with predetermined exercises. One study also showed that although some patients feel empowered and safe in the iCBT format, others experience isolation and feel burdened by the limited therapist support [28].

Blended CBT

Blended CBT (bCBT) might help bridge this gap between web-based and face-to-face treatment. In bCBT, patients receive a combination of web-based and face-to-face therapy. This may help therapists and patients adhere to treatment protocols while allowing for a different kind of flexibility during the physical consultations compared with the structured, guided self-help programs [29]. Face-to-face consultations in bCBT can alleviate the feelings of isolation that patients may experience in guided self-help, and it can help make the structured treatment content of CBT relatable and adaptable to the patient's symptom profile [30]. However, no study has yet investigated the impact of sleep

disturbance symptoms on digital interventions for MDD, including bCBT, compared with face-to-face treatments. As digital interventions are increasingly being offered to patients with MDD to augment the reach of evidence-based treatment, it is relevant to examine the association between the severity of sleep disturbance symptoms and treatment response to bCBT.

Objectives

The objectives of this study are to investigate (1) whether baseline sleep disturbance symptoms have a negative impact on treatment outcomes for depression independent of treatment modality and (2) whether sleep disturbance symptoms impact bCBT and treatment-as-usual (TAU) in routine care equally.

On the basis of previous research, we hypothesized a priori that sleep disturbance symptoms would be negatively associated with treatment outcome independent of treatment modality.

Methods

Overview

This study was a substudy of the E-COMPARED (European Comparative Effectiveness Research on Blended Depression Treatment Versus Treatment-as-Usual; described in more detail in a study by Kleiboer et al [31]).

The trial was registered in clinical trial databases (ClinicalTrials.gov NCT02542891 [France], NCT02389660 [Poland], NCT02361684 [Spain], NCT02449447 [Sweden], and NCT02410616 [Switzerland]; other clinical databases: German Clinical Trials Register DRKS00006866 [Germany]; Netherlands Trials Register NTR4962 [the Netherlands], International Standard Randomised Controlled Trial Number registry ISRCTN12388725 [United Kingdom], and ClinicalTrials.gov NCT02796573 [Denmark]) and was conducted based on CONSORT (Consolidated Standards of Reporting Trials) [32].

Study Design and Setting

The E-COMPARED trial was conducted as a 2-arm, parallel randomized controlled, noninferiority trial in nine European countries: France, Germany, the Netherlands, Poland, Spain, Sweden, Switzerland, the United Kingdom, and Denmark.

The trial was conducted in routine primary care facilities (sites: Germany, Poland, Spain, Sweden, and the United Kingdom) or specialized mental health care facilities (sites: France, the Netherlands, Switzerland, and Denmark).

Eligibility Criteria

To be included in the study, patients had to (1) be aged ≥ 18 years, (2) fulfill Diagnostic and Statistical Manual of Mental Disorders-IV (DSM-IV) diagnostic criteria for MDD, and (3) exhibit minimal to severe symptoms of depression based on a score of ≥ 5 on a screening with the Patient Health Questionnaire-9 (PHQ-9; described in more detail in the *Assessment Instruments* section).

The exclusion criteria were as follows: (1) currently being at high risk for suicide (2) having fulfilled the DSM-IV diagnostic criteria for substance dependence, psychotic illness, bipolar

affective disorder, or obsessive–compulsive disorder; (3) receiving psychological treatment at the time of enrollment for depression in primary or specialized mental health care facilities; (4) being unable to comprehend the spoken and written language of the country where the study was conducted (eg, English in the United Kingdom); (5) not having access to a computer with a fast internet connection; and (6) not having or not willing to carry a smartphone that is compatible with the mobile component of the intervention being offered.

Randomization and Blinding

Eligible participants were randomized by a team of independent researchers (the randomization team) affiliated with the principal investigator organization (Vrije Universiteit Amsterdam) into two arms: TAU or bCBT. Randomization was performed at the individual level, stratified by country. The randomization team created the allocation scheme with a computerized random number generator (Random Allocation Software developed by Mahmood Saghaei) at an allocation ratio of 1:1. Block randomization with variable block sizes that varied between 8 and 14 allocations per block was applied.

None of the investigators or clinicians were aware of the randomization scheme, but blinding for treatment allocation was not possible owing to the nature of the treatment. Nonetheless, the outcome assessors were blinded.

Treatment Arms

TAU was defined as the routine care that patients received when diagnosed with depression in the specific country and treatment setting where they were recruited. Thus, TAU varied between

countries, between treatment settings, and among patients and included pharmacological treatment, psychotherapy, a combination of both, or none of the above (see Table 1).

The bCBT combined individual face-to-face CBT with CBT delivered through an internet-based treatment platform with mobile phone components (integrated either in the treatment platform or as a separate system). The core components of the bCBT treatment were the following: (1) psychoeducation, (2) cognitive restructuring, (3) behavioral activation, and (4) relapse prevention. These were delivered over 10 to 20 sessions. The ratio between the number of face-to-face sessions and the number of web-based modules varied according to practice in the participating countries. However, a minimum of 33% of the sessions were face-to-face, and a minimum of 33% of the sessions were provided through the web. In addition to the core CBT components, other components such as mindfulness, coping skills training, or problem solving could also be included (insomnia management was not included in this trial). However, they were not allowed to take up more than a quarter of the total treatment (no more than approximately 25% of the face-to-face and web-based sessions combined). This was done to prevent excessive heterogeneity in the treatment programs.

bCBT was provided by CBT therapists who received training on how to deliver the treatment. CBT therapists were (1) licensed CBT therapists, (2) CBT therapists in training working under the supervision of an experienced licensed CBT therapist in specialized mental health care facility, (3) licensed psychologists, or (4) psychologists in training working under the supervision of a licensed psychologist with a CBT orientation in primary care.

Table 1. Overview of blended cognitive behavioral therapy (bCBT) and treatment-as-usual (TAU) applied in each country.

Country	Platform for bCBT	Duration (weeks)	Web-based modules	FTF ^a sessions	Sequencing	TAU
Germany	Moodbuster	10-13	10	5	Alternate	TAU from GP ^b
Sweden	Iterapi	10	6	4	Alternate	TAU from GP ^c
Netherlands	Moodbuster	20	10	9	Alternate	FTF TAU ^d
United Kingdom	Moodbuster	11	5	6	Alternate	FTF CBT
Spain	Smiling is fun ^e	10	8	3	1-4-1-4-1 ^f	TAU from GP
France	Moodbuster	16	8	8	Alternate	FTF CBT
Switzerland	Deprexis ^g	18	9	9	Alternate	FTF CBT
Poland	Moodbuster	6-10	6	7	Alternate	FTF CBT
Denmark	NoDep ^h	12	6-8	6	Alternate	FTF CBT

^aFTF: face-to-face.

^bGP: general practitioner.

^cSweden also included psychotherapy clinics and student mental health care facilities. However, these are at the same level of care as GP.

^dPsychotherapy (cognitive behavioral therapy, interpersonal psychotherapy, or supportive therapy), antidepressant medication, running therapy, or a combination of these.

^eAdditional module on coping skills.

^fThe sequence was as follows: 1 FTF session, 4 web-based modules, 1 FTF session, 4 web-based modules, and 1 FTF session.

^gAdditional modules on mindfulness, interpersonal skills, positive psychology, emotion-focused therapy, and childhood experiences.

^hTwo additional modules on restructuring of beliefs and management of rumination that clinicians could add to the web-based sessions if deemed necessary.

As can be seen in [Table 1](#), none of the bCBT protocols included sleep disturbance symptoms interventions. However, we do not know to what extent sleep disturbance symptoms were included in the TAU protocol.

Measures

Sociodemographic factors, MDD diagnoses, severity of depressive symptoms, and psychiatric comorbidities were measured at baseline. The clinical outcome measures used in this study were severity of depressive symptoms assessed at baseline and at the 3-, 6-, and 12-month follow-up and MDD diagnosis assessed at baseline and at the 12-month follow-up.

Assessment Instruments

MDD Severity

The PHQ-9 was used to assess the severity of depression. It includes symptom domains of MDD based on DSM-IV (ie, sleep disturbance, sad mood, appetite and weight, concentration, self-criticism, suicidal ideation, interest, energy and fatigue, and psychomotor agitation and retardation). It consists of 9 items, each scored on a 0 to 3 scale, with the total score ranging from 0 to 27 and higher scores indicating more severe depression [33]. The PHQ-9 is frequently used in clinical trials to assess treatment outcomes and can be used in different patient populations such as those who receive primary care and specialized mental health care [34]. The PHQ-9 was used to screen for the presence of at least a minimum level of MDD severity (cutoff ≥ 5). Furthermore, the PHQ-9 total score was applied as the primary outcome. For this study, the 12-month follow-up was used as the primary end point. The 3- and 6-month follow-ups were the secondary end points.

MDD Diagnosis

The Mini-International Neuropsychiatric Interview (MINI) for DSM-IV is a structured interview probing the 17 most common psychiatric diagnoses using dichotomous questions requiring a yes or no response [35]. The MINI was used to establish MDD diagnoses as well as to screen patients for eligibility and detect comorbid psychiatric diagnoses. The MINI was also used to evaluate whether the patients still fulfilled the diagnostic criteria for an MDD diagnoses at the 12-month follow-up and was applied as a secondary outcome and secondary end point.

Assessment of Sleep Disturbance Symptoms

A total of four sleep disturbance items from the Quick Inventory of Depressive Symptomatology (QIDS)–Self-Report were used to measure sleep disturbance symptoms [36]: (1) difficulty falling asleep, (2) awake during the night, (3) waking up too early, and (4) sleeping too much. Each item was scored from 0 to 3, with higher scores indicating greater symptom severity. This approach to measure sleep disturbance symptoms has been validated previously for the 3 insomnia items [37]. However, the middle-answer categories (scores of 1 and 2) did not perform as well for items 2 and 3 [37]. Therefore, on the basis of previous research, a score ≥ 2 on an item was chosen to indicate the presence of that specific type of sleep disturbance [38]. A combination of any (or none) of the 4 types of sleep disturbance symptoms was possible.

Analysis Plan

Characteristics of the sample at baseline were described using descriptive statistics and compared across groups using 2-tailed *t* tests for continuous variables and chi-square tests for categorical variables. If the continuous variables violated the assumption of normality, equivalent nonparametric tests were used (Wilcoxon signed rank test).

The primary analysis was performed using a linear mixed effects model with random intercept [39,40] with sites added as a random variable. The outcome variable was the difference in score on the PHQ-9 between baseline and at 12-month follow-up. The baseline sum score of the first 4 items on the QIDS–Self-Report pertaining to sleep disturbance symptoms was used as a predictor variable and was controlled for the baseline level of depression severity on the PHQ-9. To test for an effect of treatment modality (bCBT vs TAU), an interaction term between the sleep disturbance symptoms and treatment modality was added. A sensitivity analysis excluding item 3 in the PHQ-9 (pertaining to sleep disturbance) was also performed as this item could potentially confound the analyses.

Secondary analyses were also conducted with linear mixed effects models applying the same predictor and outcome variables as well as an interaction term, but at 3- and 6-month follow-up. In addition, an analysis using the same predictor variable and the presence versus absence of MDD diagnoses at 12-month follow-up as the outcome was conducted. This analysis was performed using a logistic mixed effects model.

All models were adjusted for the baseline level of depression on the PHQ-9 and sociodemographic variables (gender, marital status, highest education, and age).

All analyses followed the intention-to-treat principles according to CONSORT guidelines [41]. Missing values were handled by using all available data in mixed effects models. A double-sided significance level of .05 was applied. All calculations were performed using R version 3.6.3 [42]. Linear mixed effects models were fitted using the lme4 package [40].

Ethical Considerations

Ethical approval for the trials was obtained locally in each country (Denmark: De Videnskabetiske Komitéer for Region Syddanmark; S-20150150; France: Comité de protection des personnes, Ile de France V; 15033-n° 2015-A00565-44; Germany: Ethik Kommission DGPsychologie, Universität Trier; MB 102014; The Netherlands: METC VUMC; 2015.078; Poland: Komisja ds. Etyki Badan Naukowych; 10/2014; Spain: Comision Deontologica/Comite Etico de Investigacion en Humanos; H1414775276823; Sweden: Regionala etikprovningssamnden; 2014/428-31; Switzerland: Kantonale Ethikkommission Bern; 001/2015; United Kingdom: NRES Committee London-Camden and King's Cross; 15/LO/0511). All participants provided written informed consent beforehand.

Results

Characteristics of the Sample

The sample consisted of 943 adult patients with MDD recruited from routine care across the 9 participating European countries.

The mean age was 39 (SD 13) years, with mainly women included (644/943, 68.3%). The majority were either married or living together (517/943, 54.8%) and were well educated. At baseline, the sample, on average, presented with a moderate to severe degree of depression (mean PHQ-9 score 15.35, SD

4.77). There were no differences between the groups on any sociodemographic measures or on depressive or sleep symptomatology at baseline. The characteristics of the participants are summarized in [Table 2](#).

Table 2. Comparison of characteristics across treatment modalities (N=943).

Characteristic	Total sample (N=943)	TAU ^a (n=467)	bCBT ^b (n=476)	P value ^c
Age (years), mean (SD)	38.96 (13.09)	38.71 (13.08)	39.21 (13.1)	.56
Gender				
Female, mean (SD)	644 (68.3)	326 (69.8)	318 (66.8)	.36
Marital status , n (%)				.27
Single	314 (33.3)	155 (33.2)	159 (33.4)	
Divorced	103 (10.9)	43 (9.2)	60 (12.6)	
Widowed	9 (0.9)	6 (1.3)	3 (0.6)	
Living together	206 (21.8)	111 (23.8)	95 (19.9)	
Married	311 (32.9)	152 (32.5)	159 (33.4)	
Highest level of education , n (%)				.92
Low	146 (15.5)	74 (15.8)	72 (15.1)	
Middle	349 (37)	170 (36.4)	179 (37.6)	
High	447 (47.4)	222 (47.5)	225 (47.3)	
Trial site , n (%)				.99
Germany	173 (18.3)	87 (18.6)	86 (18.1)	
Sweden	141 (14.9)	68 (14.6)	73 (15.3)	
Netherlands	102 (10.8)	49 (10.5)	53 (11.1)	
United Kingdom	101 (10.7)	52 (11.1)	49 (10.3)	
Spain	127 (13.5)	63 (13.5)	64 (13.4)	
France	105 (11.1)	54 (11.6)	51 (10.7)	
Switzerland	50 (5.3)	24 (5.1)	26 (5.5)	
Poland	84 (8.9)	42 (8.9)	42 (8.8)	
Denmark	60 (6.4)	28 (5.9)	32 (6.7)	
Psychometrics				
PHQ-9 ^d , mean (SD)	15.36 (4.78)	15.38 (4.66)	15.34 (4.9)	.88
QIDS ^e sum of sleep scores (items 1-4), mean (SD)	4.79 (2.42)	4.71 (2.31)	4.87 (2.52)	.31
Insomnia items , n (%)				
Difficulty falling asleep	453 (48)	225 (48.2)	228 (47.9)	.99
Awake during the night	512 (54.3)	251 (53.7)	261 (54.8)	.76
Waking up too early	321 (34)	149 (31.9)	172 (36.1)	.18
Hypersomnia item , n (%)				
Sleeping too much	135 (14.3)	65 (13.9)	70 (14.7)	.77

^aTAU: treatment-as-usual.

^bbCBT: blended cognitive behavioral therapy.

^cA 2-tailed *t* test was performed for continuous variables (age, PHQ-9, QIDS). Chi-square test was performed for categorical variables (gender, marital status, education, trial site and, insomnia and hypersomnia prevalence).

^dPHQ-9: Patient Health Questionnaire-9.

^eQIDS: Quick Inventory of Depressive Symptomatology.

Observed Results

In this study's sample of 943 patients, 764 (81%) reported at least one symptom of sleep disturbance symptoms at baseline. The most common symptom of sleep disturbance upon entering the study was midnocturnal insomnia (512/943, 54.3%), which was reported by more than half of the participants. Sleep onset insomnia (453/943, 48%) followed in close succession.

Hypersomnia was markedly less prevalent than the rest (135/943, 14.3%).

As measured by the sum of the first 4 items in the QIDS pertaining to sleep, the level of sleep disturbance reduced over time, as seen in Table 3. In Figure 1, the levels of severity of depressive and sleep disorder symptoms are presented for the total sample and for each treatment condition.

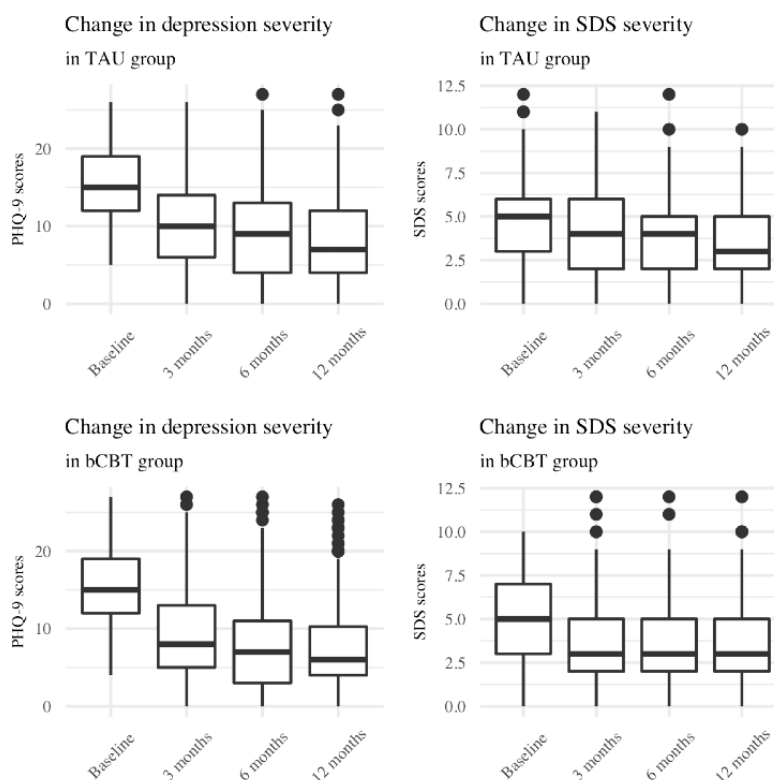
Table 3. Observed means of sleep disturbance symptom severity.

Time point	Total sample, mean (SD)	TAU ^a , mean (SD)	bCBT ^b , mean (SD)
Baseline	4.79 (2.42)	4.87 (2.52)	4.77 (2.31)
3 months	3.89 (2.43)	3.83 (2.49)	3.94 (2.37)
6 months	3.61 (2.38)	3.52 (2.42)	3.69 (2.33)
12 months	3.28 (2.39)	3.27 (2.46)	3.30 (2.33)

^aTAU: treatment-as-usual.

^bbCBT: blended cognitive behavioral therapy.

Figure 1. Changes in symptoms severity of depression and sleep disturbance symptoms from baseline to 12-month follow-up. bCBT: blended cognitive behavioral therapy; PHQ-9: Patient Health Questionnaire-9; TAU: treatment-as-usual.



Primary Analysis

In the total sample, the level of sleep disturbance symptoms did not significantly predict the change in depression symptom severity from baseline to the 12-month follow-up ($\beta=.16$, 95% CI -0.04 to 0.36). However, there was a difference between the groups, with the effect of baseline sleep disturbance symptoms being nonsignificant in the TAU group ($\beta=-.23$, 95% CI -0.50 to 0.05), but being significant for the bCBT group ($\beta=.49$, 95% CI 0.22 - 0.76). When including an interaction term with group,

the interaction was significant ($\beta=.59$, 95% CI 0.23 - 0.94), indicating a significant difference in how the change at 12 months was affected by baseline sleep disturbance symptoms between the groups. In a sensitivity analysis, we removed item 3 in the PHQ-9 (pertaining to sleep disturbances), which led to similar results. Therefore, we decided to continue using the full PHQ-9, given that it is the validated version and for the purpose of comparability. The results of the models are presented in Table 4.

Table 4. Associations between change in depression score and sleep score.

Outcome	All participants (N=943)			TAU ^a (n=467)			bCBT ^b (n=476)		
	Regression coefficient			Regression coefficient			Regression coefficient		
	n (%)	β^c	Adjusted β^d (95% CI)	n (%)	β^c	Adjusted β^d (95% CI)	n (%)	β^c	Adjusted β^d (95% CI)
Primary outcome									
Change in depression score at 12-months follow-up: sleep score at baseline, per 1 unit increase	558 (59.2)	.21	.16 (-.04 to .36)	274 (58.7)	.01	-.23 (-.50 to .05)	284 (59.7)	.35	.49 (.22 to .76)
Difference between treatment conditions: SDS ^e x condition	N/A ^f	.59	.59 (.23 to .94)	N/A	N/A	N/A	N/A	N/A	N/A
Secondary outcomes									
Change in depression score at 3-months follow-up: sleep score at baseline, per 1 unit increase	748 (79.3)	.09	.06 (-.11 to .23)	379 (81.2)	.12	.07 (-.17 to .31)	369 (77.5)	-.07	.07 (-.16 to .32)
Change in depression score at 6-months follow-up: sleep score at baseline, per 1 unit increase	657 (69.7)	.15	.09 (-.10 to .28)	319 (68.3)	.14	.03 (-.25 to .32)	338 (71)	.16	.12 (-.14 to .37)

^aTAU: treatment-as-usual.

^bbCBT: blended cognitive behavioral therapy.

^cAdjusted for depression score at baseline.

^dAdjusted for gender, age, marital status, highest education, and depression score at baseline.

^eSDS: sleep disturbance symptoms.

^fN/A: not applicable.

Secondary Analyses

In line with the primary finding, baseline sleep disturbance symptoms severity did not significantly predict the presence of MDD diagnoses for the total sample as assessed by the MINI MDD module at 12-month follow-up (odds ratio [OR] 1.05, 95% CI 0.96-1.16). However, again, the effect was significant for the bCBT group (OR 1.18, 95% CI 1.02-1.38) and not for the TAU group (OR 0.94, 95% CI 0.82-1.08).

Owing to the diversity in interventions offered in the comparison group (TAU), we performed the same analyses for a subset of sites where the patients received face-to-face CBT as the comparator (United Kingdom, France, Switzerland, Poland, and Denmark). As was the case for all sites, the results of the subgroup were not significant for the total sample ($\beta=.27$, 95% CI -0.09 to 0.63), and only the bCBT group showed a significant effect of sleep disturbance symptoms ($\beta=.69$, 95% CI 0.19 - 1.19). However, the interaction term with group was nonsignificant ($\beta=.56$, 95% CI -0.07 to 1.22). This could be owing to the subgroup including fewer participants, thus widening the CI. Analyses of the presence of depression according to the MINI did not result in any significant results, which again may be caused by the smaller sample size.

The effect of sleep disturbance symptoms on the change in the severity of depressive symptoms was nonsignificant from baseline to the 3- and 6-month follow-up for both treatment

formats, suggesting that sleep disturbance symptoms may not have an effect on short-term treatment gains.

Of the 4 items on sleep disturbance symptoms in the QIDS, 3 pertained to insomnia and 1 pertained to hypersomnia (see Table 2). Exploratory analyses were performed to examine whether there was a difference between the effect of hypersomnia and insomnia. When performing the same analysis as the primary analysis but excluding the hypersomnia item, it did not change the results. Furthermore, hypersomnia alone did not predict the treatment effect.

Discussion

Principal Findings

The objectives of this study were to investigate (1) whether baseline sleep disturbance symptoms have a negative impact on treatment outcomes independent of treatment modality and (2) whether sleep disturbance symptoms impact bCBT and TAU in routine care equally.

Overall, sleep disturbance symptoms did not have an impact on the treatment response. However, when comparing the treatment arms, a difference was observed in the effect of sleep disturbance symptoms on treatment outcome between bCBT and TAU. Higher baseline sleep disturbance symptoms indicated a decreased treatment response in the bCBT group, whereas TAU was unaffected. Interestingly, the effect of baseline sleep disturbance symptoms was only present at the 12-month

follow-up. This finding points to a long-term effect of sleep disturbance symptoms on the course of depression following bCBT, even though treatment initially seems unaffected. Despite the effect being small, it was supported by the fact that sleep disturbance symptoms also significantly affected the risk of being diagnosed with MDD at the 12-month follow-up in the bCBT group.

However, in a secondary analysis of a subgroup of sites, we compared bCBT only with face-to-face CBT. This analysis showed that although sleep disturbance symptoms still significantly impacted bCBT, there was no significant difference between treatment arms in this subgroup. Not surprisingly, this could indicate that the face-to-face and blended CBT conditions were more similar in content and outcome compared with the remaining TAU formats. Perhaps the less structured TAU conditions were slightly more attentive to the presence of sleep disturbance symptoms, indicating a shortcoming in common CBT protocols for MDD.

This study found results that support previous research and results that contradict previous findings. In line with previous research and our hypothesis, bCBT was affected by sleep disturbance symptoms. However, contrary to this, TAU and the face-to-face CBT subgroup were unaffected by sleep disturbance symptoms.

The finding that sleep disturbance symptoms have a negative impact on bCBT represents an important addition to the existing literature on sleep disturbance symptoms and MDD treatment. It is in accordance with previous research showing an increased risk for MDD nonremission following face-to-face treatment [11,12,14], as well as studies showing that insomnia symptoms decrease treatment response in iCBT [23,24]. Hence, the results of this study extend the findings from previous research by confirming the association between sleep disturbance symptoms and treatment response for MDD in a new treatment format, bCBT.

Furthermore, the analyses showed that a higher endorsement of sleep disturbance items at baseline was associated with a lower treatment response at 12-month follow-up. This corroborates the findings of Pigeon et al [13] and Troxel et al [14]. However, although these authors only used a dichotomous measure of MDD (remission vs nonremission), the analyses of this study expanded this by showing an association between the severity of sleep disturbance symptoms and the degree of symptom alleviation at 12-month follow-up.

Although the long-term effect of baseline sleep disturbance symptoms on bCBT for MDD is in line with previous research [13], the finding of this study that treatment response at the 3- and 6-month follow-up was unaffected by sleep disturbance symptoms is not consistent with previous findings. Different explanations may apply. First, none of the previous studies investigated bCBT. This finding may be uniquely associated with bCBT. However, there are no clear indications as to why this should be the case. Second, neither MDD nor sleep disturbance symptoms has been measured uniformly in previous research. Notably, this study used sleep disturbance items from a depression scale (QIDS) to measure sleep disturbance symptoms. Although this method has been validated [37,43],

it is not the most reliable measure of sleep disturbance symptoms. In the study by Troxel et al [14], subjectively measured sleep disturbance symptoms did not predict treatment outcome alone, but it did when coupled with an objective indicator. Another notable difference is the outcome measures used across these studies. As mentioned, previous research has generally focused on dichotomous outcomes [12-14], whereas this study used the PHQ-9 to measure the size of the treatment effect. If the MINI had been applied at the 3- and 6-month follow-up as well in this study, it would have been interesting to compare the dichotomous findings regarding nonremission at these time points. These differences might affect the comparability with the studies and therefore the findings of this study should not discount previous research.

The missing effect of sleep disturbance symptoms on TAU is also noteworthy. It might indicate that TAU practitioners were more attentive to sleep disturbance symptoms and implemented an intervention for sleep disturbance symptoms when necessary. However, this is still contrary to the study by Pigeon et al [13], who found a significant effect of insomnia on MDD treatment in routine care. Therefore, it may be worth noting that the study by Pigeon et al [13] is relatively old. It could be hypothesized that owing to the increased attention to sleep disturbance symptoms in the past years [8,10], routine care practitioners may treat patients reporting sleep disturbance symptoms differently today. This may help explain why TAU was unaffected by sleep disturbance symptoms in this study. However, it is highly speculative.

Strengths and Limitations

Some strengths of this study are worth noting. Primarily, this study used a randomized and longitudinal design with multiple measurement points. Using this design, we can determine the extent to which bCBT compares with established effective treatments. Furthermore, we can determine the extent to which treatment effects are sustained following treatment.

Second, this study used a large, heterogeneous sample spanning 9 different European countries, lending the study great ecological validity. Furthermore, the comparison between TAU and bCBT provides clinically useful results, as TAU reflects the everyday treatments delivered by clinicians rather than the standardized treatments of a research trial. This part of the design could also be framed as a weakness because an efficacy study using only standardized protocols can be easier to interpret. However, this would have been at the cost of the finding's generalizability [44], which was highly important in the E-COMPARED study [31].

Some limitations of the analyses of this study should also be considered. Primarily, the E-COMPARED study was not geared to investigate a research question regarding sleep disturbance symptoms, which is highlighted by the lack of a dedicated insomnia measure. However, as indicated by Manber et al [37], the items from the QIDS perform satisfactorily in discriminating the presence of sleep disturbance symptoms. Furthermore, we decided to pool hypersomnia and insomnia items to establish an aggregated score for sleep disturbance symptoms. Given the higher prevalence of insomnia, as was also evident in the sample of this study, this approach might be considered questionable.

Focusing solely on insomnia symptoms would provide a clearer picture of what sleep disturbance symptoms entail. However, given that hypersomnia and insomnia symptoms may co-occur and have been associated with more severe depression in adolescents [18,19], capturing the full phenomenon seemed prudent.

This study provides important findings regarding the use of bCBT to treat MDD with marked sleep disturbance symptoms. When the treatment format moves from face-to-face to internet-based, the potential for broader dissemination generally comes at the cost of treatment flexibility. The analyses of this study indicated that sleep disturbance symptoms might affect the less flexible treatment format, bCBT, more than it affects TAU. However, given that this was a secondary analysis of a study that was not designed to test this hypothesis, the results should be interpreted cautiously. It will be important for future research to test whether these findings are supported when using dedicated sleep disturbance measures. Nonetheless, it would also be of interest to see if treatment outcomes in bCBT can be improved by using web-based add-on sessions for sleep disturbance symptoms. Previous research has shown that iCBT

can effectively treat sleep disturbance symptoms [25,45-48], and there is a call for research on sequential or concomitant treatment of MDD and sleep disturbance symptoms [49]. Finally, it is of interest to investigate whether these findings extend to other forms of diagnostic complexity. Future studies should compare bCBT with face-to-face therapy for patients with MDD with other common comorbidities such as alcohol use or anxiety disorders.

Conclusions

Baseline sleep disturbance symptoms may have a negative impact on long-term treatment outcomes in bCBT for MDD. TAU seems to be unaffected by the severity of the baseline sleep disturbance symptoms. This suggests that extra attention to comorbidity such as sleep disturbance symptoms, which was investigated here, might be important when treating depression with bCBT in routine care.

Future research should investigate whether there is an improved treatment outcome in bCBT for MDD with comorbid sleep disturbance symptoms when using add-on modules targeting sleep disturbance symptoms.

Acknowledgments

The authors thank the E-COMPARED (European Comparative Effectiveness Research on Blended Depression Treatment Versus Treatment-as-Usual) consortium for giving us the opportunity to conduct this secondary analysis on their trial data. In addition, the authors thank Innovation Fund Denmark for funding the Danish trial in the ENTER project, which delivered data to the E-COMPARED project.

Conflicts of Interest

None declared.

References

1. Ferrari AJ, Somerville AJ, Baxter AJ, Norman R, Patten SB, Vos T, et al. Global variation in the prevalence and incidence of major depressive disorder: a systematic review of the epidemiological literature. *Psychol Med* 2013 Mar;43(3):471-481. [doi: [10.1017/S0033291712001511](https://doi.org/10.1017/S0033291712001511)] [Medline: [22831756](https://pubmed.ncbi.nlm.nih.gov/22831756/)]
2. Tylee A, Gastpar M, Lépine JP, Mendlewicz J. DEPRES II (Depression Research in European Society II): a patient survey of the symptoms, disability and current management of depression in the community. DEPRES Steering Committee. *Int Clin Psychopharmacol* 1999 May;14(3):139-151. [doi: [10.1097/00004850-199905002-00001](https://doi.org/10.1097/00004850-199905002-00001)] [Medline: [10435767](https://pubmed.ncbi.nlm.nih.gov/10435767/)]
3. Findings from the Global Burden of Disease Study. Seattle, WA: Institute for Health Metrics and Evaluation; 2018.
4. Olesen J, Gustavsson A, Svensson M, Wittchen H, Jönsson B, CDBE2010 study group, European Brain Council. The economic cost of brain disorders in Europe. *Eur J Neurol* 2012 Jan;19(1):155-162. [doi: [10.1111/j.1468-1331.2011.03590.x](https://doi.org/10.1111/j.1468-1331.2011.03590.x)] [Medline: [22175760](https://pubmed.ncbi.nlm.nih.gov/22175760/)]
5. Cuijpers P. Four decades of outcome research on psychotherapies for adult depression: an overview of a series of meta-analyses. *Canadian Psychology/Psychologie canadienne* 2017;58(1):7-19. [doi: [10.1037/cap0000096](https://doi.org/10.1037/cap0000096)]
6. Thornicroft G, Chatterji S, Evans-Lacko S, Gruber M, Sampson N, Aguilar-Gaxiola S, et al. Undertreatment of people with major depressive disorder in 21 countries. *Br J Psychiatry* 2017 Feb;210(2):119-124 [FREE Full text] [doi: [10.1192/bjp.bp.116.188078](https://doi.org/10.1192/bjp.bp.116.188078)] [Medline: [27908899](https://pubmed.ncbi.nlm.nih.gov/27908899/)]
7. Barlow DH, Bullis JR, Comer JS, Ametaj AA. Evidence-based psychological treatments: an update and a way forward. *Annu Rev Clin Psychol* 2013;9:1-27. [doi: [10.1146/annurev-clinpsy-050212-185629](https://doi.org/10.1146/annurev-clinpsy-050212-185629)] [Medline: [23245338](https://pubmed.ncbi.nlm.nih.gov/23245338/)]
8. Franzen PL, Buysse DJ. Sleep disturbances and depression: risk relationships for subsequent depression and therapeutic implications. *Dialogues Clin Neurosci* 2008;10(4):473-481 [FREE Full text] [Medline: [19170404](https://pubmed.ncbi.nlm.nih.gov/19170404/)]
9. Diagnostic and statistical manual of mental disorders (DSM-5). Washington, DC: American Psychiatric Association; 2013.
10. Riemann D, Krone LB, Wulff K, Nissen C. Sleep, insomnia, and depression. *Neuropsychopharmacology* 2020 Jan;45(1):74-89 [FREE Full text] [doi: [10.1038/s41386-019-0411-y](https://doi.org/10.1038/s41386-019-0411-y)] [Medline: [31071719](https://pubmed.ncbi.nlm.nih.gov/31071719/)]

11. Bei B, Asarnow LD, Krystal A, Edinger JD, Buysse DJ, Manber R. Treating insomnia in depression: insomnia related factors predict long-term depression trajectories. *J Consult Clin Psychol* 2018 Mar;86(3):282-293 [FREE Full text] [doi: [10.1037/ccp0000282](https://doi.org/10.1037/ccp0000282)] [Medline: [29504795](https://pubmed.ncbi.nlm.nih.gov/29504795/)]
12. Buysse DJ, Tu XM, Cherry CR, Begley AE, Kowalski J, Kupfer DJ, et al. Pretreatment REM sleep and subjective sleep quality distinguish depressed psychotherapy remitters and nonremitters. *Biological Psychiatry* 1999 Jan;45(2):205-213. [doi: [10.1016/s0006-3223\(98\)00198-x](https://doi.org/10.1016/s0006-3223(98)00198-x)]
13. Pigeon W, Hegel M, Unützer J, Fan MY, Sateia MJ, Lyness JM, et al. Is insomnia a perpetuating factor for late-life depression in the IMPACT cohort? *Sleep* 2008 Apr;31(4):481-488 [FREE Full text] [doi: [10.1093/sleep/31.4.481](https://doi.org/10.1093/sleep/31.4.481)] [Medline: [18457235](https://pubmed.ncbi.nlm.nih.gov/18457235/)]
14. Troxel WM, Kupfer DJ, Reynolds III CF, Frank E, Thase ME, Miewald JM, et al. Insomnia and objectively measured sleep disturbances predict treatment outcome in depressed patients treated with psychotherapy or psychotherapy-pharmacotherapy combinations. *J Clin Psychiatry* 2011 Nov 29;73(04):478-485. [doi: [10.4088/jcp.11m07184](https://doi.org/10.4088/jcp.11m07184)]
15. Carney CE, Segal ZV, Edinger JD, Krystal AD. A comparison of rates of residual insomnia symptoms following pharmacotherapy or cognitive-behavioral therapy for major depressive disorder. *J Clin Psychiatry* 2007 Feb;68(2):254-260. [doi: [10.4088/jcp.v68n0211](https://doi.org/10.4088/jcp.v68n0211)] [Medline: [17335324](https://pubmed.ncbi.nlm.nih.gov/17335324/)]
16. Karp JF, Buysse DJ, Houck PR, Cherry C, Kupfer DJ, Frank E. Relationship of variability in residual symptoms with recurrence of major depressive disorder during maintenance treatment. *Am J Psychiatry* 2004 Oct;161(10):1877-1884. [doi: [10.1176/ajp.161.10.1877](https://doi.org/10.1176/ajp.161.10.1877)] [Medline: [15465986](https://pubmed.ncbi.nlm.nih.gov/15465986/)]
17. Pigeon WR, May PE, Perlis ML, Ward EA, Lu N, Talbot NL. The effect of interpersonal psychotherapy for depression on insomnia symptoms in a cohort of women with sexual abuse histories. *J Trauma Stress* 2009 Dec;22(6):634-638 [FREE Full text] [doi: [10.1002/jts.20456](https://doi.org/10.1002/jts.20456)] [Medline: [19885874](https://pubmed.ncbi.nlm.nih.gov/19885874/)]
18. Soehner AM, Kaplan KA, Harvey AG. Prevalence and clinical correlates of co-occurring insomnia and hypersomnia symptoms in depression. *J Affect Disord* 2014;167:93-97 [FREE Full text] [doi: [10.1016/j.jad.2014.05.060](https://doi.org/10.1016/j.jad.2014.05.060)] [Medline: [24953480](https://pubmed.ncbi.nlm.nih.gov/24953480/)]
19. Liu X, Buysse DJ, Gentzler AL, Kiss E, Mayer L, Kapornai K, et al. Insomnia and hypersomnia associated with depressive phenomenology and comorbidity in childhood depression. *Sleep* 2007 Jan;30(1):83-90. [doi: [10.1093/sleep/30.1.83](https://doi.org/10.1093/sleep/30.1.83)] [Medline: [17310868](https://pubmed.ncbi.nlm.nih.gov/17310868/)]
20. Karyotaki E, Efthimiou O, Miguel C, Berrpohl FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry* 2021 Apr 01;78(4):361-371. [doi: [10.1001/jamapsychiatry.2020.4364](https://doi.org/10.1001/jamapsychiatry.2020.4364)] [Medline: [33471111](https://pubmed.ncbi.nlm.nih.gov/33471111/)]
21. 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](https://doi.org/10.1016/j.janxdis.2018.01.001)] [Medline: [29422409](https://pubmed.ncbi.nlm.nih.gov/29422409/)]
22. Andrade LH, Alonso J, Mneimneh Z, Wells JE, Al-Hamzawi A, Borges G, et al. Barriers to mental health treatment: results from the WHO World Mental Health surveys. *Psychol Med* 2013 Aug 09;44(6):1303-1317. [doi: [10.1017/s0033291713001943](https://doi.org/10.1017/s0033291713001943)]
23. El Alaoui S, Ljótsson B, Hedman E, Svanborg C, Kaldø V, Lindfors N. Predicting outcome in internet-based cognitive behaviour therapy for major depression: a large cohort study of adult patients in routine psychiatric care. *PLoS One* 2016;11(9):e0161191 [FREE Full text] [doi: [10.1371/journal.pone.0161191](https://doi.org/10.1371/journal.pone.0161191)] [Medline: [27618548](https://pubmed.ncbi.nlm.nih.gov/27618548/)]
24. Lindsäter E, Axelsson E, Salomonsson S, Santoft F, Ljótsson B, Åkerstedt T, et al. The mediating role of insomnia severity in internet-based cognitive behavioral therapy for chronic stress: secondary analysis of a randomized controlled trial. *Behav Res Ther* 2021 Jan;136:103782 [FREE Full text] [doi: [10.1016/j.brat.2020.103782](https://doi.org/10.1016/j.brat.2020.103782)] [Medline: [33276274](https://pubmed.ncbi.nlm.nih.gov/33276274/)]
25. Blom K, Jernelöv S, Kraepelien M, Bergdahl MO, Jungmarker K, Ankartjärn L, et al. Internet treatment addressing either insomnia or depression, for patients with both diagnoses: a randomized trial. *Sleep* 2015 Feb 01;38(2):267-277 [FREE Full text] [doi: [10.5665/sleep.4412](https://doi.org/10.5665/sleep.4412)] [Medline: [25337948](https://pubmed.ncbi.nlm.nih.gov/25337948/)]
26. Ye Y, Zhang Y, Chen J, Liu J, Li X, Liu Y, et al. Internet-based cognitive behavioral therapy for insomnia (ICBT-i) improves comorbid anxiety and depression—a meta-analysis of randomized controlled trials. *PLoS One* 2015;10(11):e0142258 [FREE Full text] [doi: [10.1371/journal.pone.0142258](https://doi.org/10.1371/journal.pone.0142258)] [Medline: [26581107](https://pubmed.ncbi.nlm.nih.gov/26581107/)]
27. Andersson G, Bergström J, Holländare F, Carlbring P, Kaldø V, Ekselius L. Internet-based self-help for depression: randomised controlled trial. *Br J Psychiatry* 2005 Nov;187:456-461. [doi: [10.1192/bjp.187.5.456](https://doi.org/10.1192/bjp.187.5.456)] [Medline: [16260822](https://pubmed.ncbi.nlm.nih.gov/16260822/)]
28. Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. *PLoS One* 2014;9(1):e84323 [FREE Full text] [doi: [10.1371/journal.pone.0084323](https://doi.org/10.1371/journal.pone.0084323)] [Medline: [24465404](https://pubmed.ncbi.nlm.nih.gov/24465404/)]
29. Kooistra LC, Ruwaard J, Wiersma JE, van Oppen P, van der Vaart R, van Gemert-Pijnen JE, et al. Development and initial evaluation of blended cognitive behavioural treatment for major depression in routine specialized mental health care. *Internet Interv* 2016 May;4:61-71 [FREE Full text] [doi: [10.1016/j.invent.2016.01.003](https://doi.org/10.1016/j.invent.2016.01.003)] [Medline: [30135791](https://pubmed.ncbi.nlm.nih.gov/30135791/)]
30. Wilhelmsen M, Lillevoll K, Risør M, Høifødt R, Johansen M, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. *BMC Psychiatry* 2013 Nov 7;13(1):296. [doi: [10.1186/1471-244x-13-296](https://doi.org/10.1186/1471-244x-13-296)]

31. Kleiboer A, Smit J, Bosmans J, Ruwaard J, Andersson G, Topooco N, et al. European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. *Trials* 2016 Aug 03;17(1):387 [FREE Full text] [doi: [10.1186/s13063-016-1511-1](https://doi.org/10.1186/s13063-016-1511-1)] [Medline: [27488181](https://pubmed.ncbi.nlm.nih.gov/27488181/)]
32. Moher D, Hopewell S, Schulz K, Montori V, Gøtzsche PC, Devereaux PJ, Consolidated Standards of Reporting Trials Group. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol* 2010 Aug;63(8):e1-37 [FREE Full text] [doi: [10.1016/j.jclinepi.2010.03.004](https://doi.org/10.1016/j.jclinepi.2010.03.004)] [Medline: [20346624](https://pubmed.ncbi.nlm.nih.gov/20346624/)]
33. 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](https://doi.org/10.1046/j.1525-1497.2001.016009606.x)] [Medline: [11556941](https://pubmed.ncbi.nlm.nih.gov/11556941/)]
34. Costantini L, Pasquarella C, Odone A, Colucci ME, Costanza A, Serafini G, et al. Screening for depression in primary care with Patient Health Questionnaire-9 (PHQ-9): a systematic review. *J Affect Disord* 2021 Jan 15;279:473-483. [doi: [10.1016/j.jad.2020.09.131](https://doi.org/10.1016/j.jad.2020.09.131)] [Medline: [33126078](https://pubmed.ncbi.nlm.nih.gov/33126078/)]
35. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J Clin Psychiatry* 1998;59 Suppl 20:22-33;quiz 34. [Medline: [9881538](https://pubmed.ncbi.nlm.nih.gov/9881538/)]
36. Rush A, Trivedi MH, Ibrahim HM, Carmody TJ, Arnow B, Klein DN, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry* 2003 Sep 01;54(5):573-583. [doi: [10.1016/s0006-3223\(02\)01866-8](https://doi.org/10.1016/s0006-3223(02)01866-8)] [Medline: [12946886](https://pubmed.ncbi.nlm.nih.gov/12946886/)]
37. Manber R, Blasey C, Arnow B, Markowitz JC, Thase ME, Rush AJ, et al. Assessing insomnia severity in depression: comparison of depression rating scales and sleep diaries. *J Psychiatr Res* 2005 Sep;39(5):481-488. [doi: [10.1016/j.jpsychires.2004.12.003](https://doi.org/10.1016/j.jpsychires.2004.12.003)] [Medline: [15992557](https://pubmed.ncbi.nlm.nih.gov/15992557/)]
38. Sunderajan P, Gaynes BN, Wisniewski SR, Miyahara S, Fava M, Akingbala F, et al. Insomnia in patients with depression: a STAR*D report. *CNS Spectr* 2010 Jun;15(6):394-404. [doi: [10.1017/s1092852900029266](https://doi.org/10.1017/s1092852900029266)] [Medline: [20625372](https://pubmed.ncbi.nlm.nih.gov/20625372/)]
39. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: keep it maximal. *J Mem Lang* 2013 Apr;68(3):255-278 [FREE Full text] [doi: [10.1016/j.jml.2012.11.001](https://doi.org/10.1016/j.jml.2012.11.001)] [Medline: [24403724](https://pubmed.ncbi.nlm.nih.gov/24403724/)]
40. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using. *J Stat Soft* 2015;67(1):1-48. [doi: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)]
41. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *Lancet* 2001 Apr 14;357(9263):1191-1194. [Medline: [11323066](https://pubmed.ncbi.nlm.nih.gov/11323066/)]
42. The R project for statistical computing. The R Foundation. URL: <https://www.R-project.org/> [accessed 2022-01-21]
43. MacGregor KL, Funderburk JS, Pigeon W, Maisto SA. Evaluation of the PHQ-9 Item 3 as a screen for sleep disturbance in primary care. *J Gen Intern Med* 2012 Mar 24;27(3):339-344 [FREE Full text] [doi: [10.1007/s11606-011-1884-5](https://doi.org/10.1007/s11606-011-1884-5)] [Medline: [21948205](https://pubmed.ncbi.nlm.nih.gov/21948205/)]
44. Streiner DL. The 2 "Es" of research: efficacy and effectiveness trials. *Can J Psychiatry* 2002 Aug;47(6):552-556. [doi: [10.1177/070674370204700607](https://doi.org/10.1177/070674370204700607)] [Medline: [12211883](https://pubmed.ncbi.nlm.nih.gov/12211883/)]
45. Kaldo V, Jernelöv S, Blom K, Ljótsson B, Brodin M, Jörgensen M, et al. Guided internet cognitive behavioral therapy for insomnia compared to a control treatment - A randomized trial. *Behav Res Ther* 2015 Aug;71:90-100 [FREE Full text] [doi: [10.1016/j.brat.2015.06.001](https://doi.org/10.1016/j.brat.2015.06.001)] [Medline: [26091917](https://pubmed.ncbi.nlm.nih.gov/26091917/)]
46. Horsch CH, Lancee J, Griffioen-Both F, Spruit S, Fitrianie S, Neerincx MA, et al. Mobile phone-delivered cognitive behavioral therapy for insomnia: a randomized waitlist controlled trial. *J Med Internet Res* 2017 Apr 11;19(4):e70 [FREE Full text] [doi: [10.2196/jmir.6524](https://doi.org/10.2196/jmir.6524)] [Medline: [28400355](https://pubmed.ncbi.nlm.nih.gov/28400355/)]
47. Rajabi Majd N, Broström A, Ulander M, Lin C, Griffiths MD, Imani V, et al. Efficacy of a theory-based cognitive behavioral technique app-based intervention for patients with insomnia: randomized controlled trial. *J Med Internet Res* 2020 Apr 01;22(4):e15841 [FREE Full text] [doi: [10.2196/15841](https://doi.org/10.2196/15841)] [Medline: [32234700](https://pubmed.ncbi.nlm.nih.gov/32234700/)]
48. 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](https://doi.org/10.2196/12686)] [Medline: [30973344](https://pubmed.ncbi.nlm.nih.gov/30973344/)]
49. Asarnow LD, Manber R. Cognitive behavioral therapy for insomnia in depression. *Sleep Med Clin* 2019 Jun;14(2):177-184 [FREE Full text] [doi: [10.1016/j.jsmc.2019.01.009](https://doi.org/10.1016/j.jsmc.2019.01.009)] [Medline: [31029185](https://pubmed.ncbi.nlm.nih.gov/31029185/)]

Abbreviations

bCBT: blended cognitive behavioral therapy

CBT: cognitive behavioral therapy

CONSORT: Consolidated Standards of Reporting Trials

DSM-IV: Diagnostic and Statistical Manual of Mental Disorders-IV

E-COMPARED: European Comparative Effectiveness Research on Blended Depression Treatment Versus Treatment-as-Usual

iCBT: internet-based cognitive behavioral therapy
MDD: major depressive disorder
MINI: Mini-International Neuropsychiatric Interview
OR: odds ratio
PHQ-9: Patient Health Questionnaire-9
QIDS: Quick Inventory of Depressive Symptomatology
TAU: treatment-as-usual

Edited by A Mavragani; submitted 21.05.21; peer-reviewed by K Matsumoto, H McCall; comments to author 30.07.21; revised version received 23.08.21; accepted 13.12.21; published 21.03.22.

Please cite as:

*Jensen ES, Ladegaard N, Mellentin AI, Ebert DD, Titzler I, Araya R, Cerga Pashoja A, Hazo JB, Holtzmann J, Cieslak R, Smoktunowicz E, Baños R, Herrero R, García-Palacios A, Botella C, Berger T, Krieger T, Holmberg TT, Topooco N, Andersson G, van Straten A, Kemmeren L, Kleiboer A, Riper H, Mathiasen K
Effect of Sleep Disturbance Symptoms on Treatment Outcome in Blended Cognitive Behavioral Therapy for Depression (E-COMPARED Study): Secondary Analysis
J Med Internet Res 2022;24(3):e30231
URL: <https://www.jmir.org/2022/3/e30231>
doi: [10.2196/30231](https://doi.org/10.2196/30231)
PMID: [35311687](https://pubmed.ncbi.nlm.nih.gov/35311687/)*

©Esben Skov Jensen, Nicolai Ladegaard, Angelina Isabella Mellentin, David Daniel Ebert, Ingrid Titzler, Ricardo Araya, Arlinda Cerga Pashoja, Jean-Baptiste Hazo, Jérôme Holtzmann, Roman Cieslak, Ewelina Smoktunowicz, Rosa Baños, Rocio Herrero, Azucena García-Palacios, Cristina Botella, Thomas Berger, Tobias Krieger, Trine Theresa Holmberg, Naira Topooco, Gerhard Andersson, Annemieke van Straten, Lise Kemmeren, Annet Kleiboer, Heleen Riper, Kim Mathiasen. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org/>), 21.03.2022. 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.

Original Paper

Internet-Delivered Cognitive Behavioral Therapy for Generalized Anxiety Disorder in Nationwide Routine Care: Effectiveness Study

Ville Ritola¹, MA; Jari Olavi Lipsanen², MA; Satu Pihlaja¹, MA, LicPs; Eero-Matti Gummerus¹, MA; Jan-Henry Stenberg¹, MA, PhD; Suoma Saarni¹, MD, PhD; Grigori Joffe¹, MD, PhD, Prof Dr

¹Department of Psychiatry, Helsinki University Hospital and University of Helsinki, Helsinki, Finland

²Department of Psychology and Logopedics, University of Helsinki, Helsinki, Finland

Corresponding Author:

Grigori Joffe, MD, PhD, Prof Dr

Department of Psychiatry

Helsinki University Hospital and University of Helsinki

P.O. Box 100

Helsinki, 00029 HUS

Finland

Phone: 358 405136500

Email: grigori.joffe@hus.fi

Abstract

Background: Therapist-supported, internet-delivered cognitive behavioral therapy (iCBT) is efficacious for generalized anxiety disorder (GAD), but few studies are yet to report its effectiveness in routine care.

Objective: In this study, we aim to examine whether a new 12-session iCBT program for GAD is effective in nationwide routine care.

Methods: We administered a specialized, clinic-delivered, therapist-supported iCBT for GAD in 1099 physician-referred patients. The program was free of charge for patients, and the completion time was not predetermined. We measured symptoms with web-based questionnaires. The primary measure of anxiety was the GAD 7-item scale (GAD-7); secondary measures were, for pathological worry, the Penn State Worry Questionnaire and, for anxiety and impairment, the Overall Anxiety Severity and Impairment Scale.

Results: Patients completed a mean 7.8 (SD 4.2; 65.1%) of 12 sessions, and 44.1% (485/1099) of patients completed all sessions. The effect size in the whole sample for GAD-7 was large (Cohen $d=0.97$, 95% CI 0.88-1.06). For completers, effect sizes were very large (Cohen $d=1.34$, 95% CI 1.25-1.53 for GAD-7; Cohen $d=1.14$, 95% CI 1.00-1.27 for Penn State Worry Questionnaire; and Cohen $d=1.23$, 95% CI 1.09-1.37 for Overall Anxiety Severity and Impairment Scale). Noncompleters also benefited from the treatment. Greater symptomatic GAD-7-measured relief was associated with more completed sessions, older age, and being referred from private or occupational care. Of the 894 patients with a baseline GAD-7 score ≥ 10 , approximately 421 (47.1%) achieved reliable recovery.

Conclusions: This nationwide, free-of-charge, therapist-supported HUS Helsinki University Hospital-iCBT for GAD was effective in routine care, but further research must establish effectiveness against other treatments and optimize the design of iCBT for GAD for different patient groups and individual patients.

(*J Med Internet Res* 2022;24(3):e29384) doi:[10.2196/29384](https://doi.org/10.2196/29384)

KEYWORDS

CBT; iCBT; cognitive behavioral therapy; routine care; generalized anxiety disorder; internet; web-based; digital health; mental health

Introduction

Background

Generalized anxiety disorder (GAD) [1] is a common but underdiagnosed and undertreated condition [2,3]. The global annual prevalence of GAD in high-income countries may be 2.3%, and this disorder is associated with increased functional impairment [4].

Pharmacotherapy and psychotherapy are considered first-line GAD treatments [5,6]. Among psychotherapies, cognitive behavioral therapy (CBT) is the most studied, with large effect sizes (Hedges $g=0.90$ [7]). However, traditional face-to-face CBT is relatively resource-consuming and limited in access.

The need to resolve the accessibility and affordability challenges of face-to-face CBT led to the development of internet-delivered CBT (iCBT), with notable work occurring in Sweden, the Netherlands, Denmark, Australia, and Canada, for example [8]. A typical iCBT intervention is location-independent and includes a program with therapeutic content and homework. It is accessible 24/7 through a web-based platform or mobile app, with or without therapist support. In therapist-supported iCBT, remote therapist support is most often asynchronous. While the intensity of therapist support in iCBT can vary, it requires far less therapist time compared with face-to-face or real-time remote psychotherapy and thus may be more cost-effective [9,10].

As demonstrated in recent meta-analyses of randomized controlled trials (RCTs) [9,11-16], therapist-supported iCBTs are efficacious in several psychiatric disorders and may be as efficacious as short-term face-to-face CBT. Therapist-supported iCBTs for depression and anxiety have higher efficacy and adherence than unguided programs do [12,17-19], which is likely due to the high level of client-therapist alliance, comparable with that in face-to-face therapies [20]. In general, iCBTs reduce symptomatic deterioration more than waiting lists (3.1% or 5.8% vs 17.4% [13,21]). For therapist-supported iCBT for GAD, efficacy has also been established for both diagnosis-specific and transdiagnostic programs [9,16].

However, RCT efficacy studies, being the gold standard for clinical evidence, have their own shortcomings. Participants are more likely to be highly motivated, have had thorough screening, and have received more intensive treatment than in routine care. Therefore, RCT-validated treatments may or may not yield similar results or level of adherence in routine care depending on whether the criteria are lax [22-24]. Addressing these limitations, routine care studies are now accepted as valid scientific evidence for their clinical effectiveness and safety [25]. Although there is strong real-world evidence for the effectiveness of therapist-supported iCBT in general [8,9,13], studies focusing on GAD are limited.

We identified 5 publications comprising 7 therapist-supported iCBT interventions focusing on GAD in routine care using 3 different programs at 2 clinics. ThisWayUp clinic in Australia performed 6 interventions comprising 2 programs [23,26-28], and Online Therapy Unit in Canada performed 1 intervention [29]. These studies mainly reported large intention-to-treat (ITT)

effect sizes (0.91-1.30), although 1 outlier study reported effects as large as 2.06 and 2.10 [28]. ThisWayUp trials reported full completion rates of 36% to 55%. Although these 3 programs demonstrated the overall feasibility of iCBT for GAD, the recruitment schemes, program length, and therapist support intensity in these trials varied widely (see the Design Comparison section in the Discussion section), and the Online Therapy Unit's trial was relatively small. More routine care studies are needed to confirm their conclusions and elucidate the optimal program design.

The HUS Helsinki University Hospital has developed and is providing nationwide, original Finnish language therapist-supported iCBT programs for several psychiatric conditions (further referred to as HUS-iCBTs), including one for GAD. Intake for the 12-session program requires a physician's referral, it is free of charge, and therapist support is provided centrally by a specialized clinic. As the combination of setting and design of the HUS-iCBT for GAD differs from those studied earlier, the effectiveness of these programs may differ.

Objective

The aim of this study is to examine the effectiveness of HUS-iCBT for GAD in routine care. We hypothesized that the intervention would have a large overall effect size and the patients' completion rate would be comparable (36%-55%) with the reported routine care studies.

Methods

Setting and Design

The HUS-iCBTs were delivered centrally by the iCBT clinic at HUS Psychiatry. The interventions were free of charge, diagnosis-specific, and therapist-supported programs. All physicians in Finland can refer their patients to therapy. The referring physicians received support from the instructions on the web. A specially trained mental health professional screened all referrals centrally. We deliberately implemented these inclusion procedures to enable virtually unlimited nationwide access for the clinical population despite the limited number of therapists.

The study was an observational, nationwide, open-label, real-world trial.

Participants

Participants were recruited from those entering HUS-iCBT for GAD between February 2016 and December 2018. To be accepted for the treatment, patients had to (1) be diagnosed with GAD (ICD-10 [1]), (2) have an email address, and (3) have an internet-based bank account or mobile ID (the means of official e-identification in Finland). Comorbidity (other than the exclusion criteria) and concomitant pharmacological and psychological treatments were allowed, as the intervention was part of routine care. Exclusion criteria were acute psychosis or mania, severe personality disorders, severe suicidality, or neurological or neuropsychiatric disorder with cognitive decline. These were screened centrally from the physician's referral before the referral was accepted. The only additional inclusion

criterion for the study was a baseline score of ≥ 8 on the GAD 7-item scale (GAD-7 [30]).

Ethical Considerations

The patients provided informed consent after the first log-in. The study protocol was approved by the ethics committee of the HUS and pertinent institutional authorities.

Intervention and Procedure

The new iCBT program for GAD was designed by an expert group to be diagnosis-specific because the existing evidence base was clearly the strongest at the time, in 2013. This new program consisted of 12 consecutive sessions and a follow-up session 3 months after treatment completion. The program was theoretically based on several models of GAD and anxiety, including aspects of the cognitive avoidance model [31], model of intolerance of uncertainty [32], metacognitive therapy [33], and Acceptance and Commitment Therapy [34], as well as social

aspects, such as assertiveness training. The sessions included text and videos, as well as educational illustrations, example stories, therapeutic exercises, and homework (Textbox 1).

After approval, the patients received an email and a letter prompting them to sign in and begin treatment. A schedule of 1 session per week was recommended, and a minimum 24-hour waiting period was enforced between sessions to encourage daily life practice. The program sent email prompts for arriving messages and, after 2 weeks of inactivity, log-in reminders. Nevertheless, no maximum completion time was required if the patient remained active in therapy.

Although HUS-iCBT for GAD was therapist-supported, several persuasive elements used in unguided iCBT programs [35] were originally included and are listed in Textbox 2. In total, 7 elements were used (or 6, if discounting tunneling as the original authors did).

Textbox 1. Content by session.

Session titles
<ul style="list-style-type: none"> • Session 1: Introduction to HUS Helsinki University Hospital–internet-delivered cognitive behavioral therapy program, cognitive behavioral therapy model, and generalized anxiety disorder • Session 2: Bodily stress response, worry and relaxation • Session 3: Worry and avoidance, intolerance of uncertainty • Session 4: Experiential Avoidance, Core Beliefs • Session 5: Negative beliefs about worrying, challenging worrying • Session 6: Positive beliefs about worrying, challenging beliefs about worrying • Session 7: Challenging worrying • Session 8: Acceptance of worries, acceptance vs submission • Session 9: Intolerance of uncertainty and reaching for perfection and certainty • Session 10: Problem solving vs worrying, solvable vs unsolvable worries • Session 11: Social skills, needs in relationships, assertiveness • Session 12: Summarizing, warning signs, plan for the future, feedback

Textbox 2. Identified persuasive design principles.

Principles and brief example
<ul style="list-style-type: none"> • Reduction: simple stepped instructions; for example, worry diary or relaxation • Tunneling: logical thematic progression • Self-monitoring: symptom graphs • Simulation: example stories with avatars • Rehearsal: web-based worry diary and relaxation training • Reminders: log-in reminders • Normative influence: normalization of common generalized anxiety disorder features

Therapist Support

The therapists in the program were clinical psychologists, psychology students, or nurses with additional therapeutic training, all working at HUS Psychiatry. Each therapist received 1 day of training on internet delivery of CBT, text-based

communication, the intervention protocol, and Good Clinical Practice. Therapists received regular group supervision and sought consultation with a senior psychotherapist at any time.

To support patients' progress, the therapists provided empathic asynchronous feedback with written messages 4 times or more

often (if the patient requested) during the therapy. If the patient was inactive for 2 weeks, the therapist pursued contact through an SMS text message or iCBT program. If no reply arrived within a week, further contact was attempted by phone and thereafter by a letter including a 2-week deadline for continuing, after which access to therapy was discontinued.

A web-based report of live data on individual therapists was created during the study to allow for supervision of compliance with the intervention protocol and to prompt support to those failing to ask for support. The report included data on the date of the last log-in and the number of patients not contacted or logged in within the last 2 weeks. The therapist time per patient was not subject to monitoring, but each therapist treated a minimum quota of patients per dedicated working hour. On average, this quota would mean 9 to 11 minutes of working time per patient per week.

Measures

Overview

Symptoms were measured using web-based questionnaires. The primary measure of anxiety was the GAD-7; secondary measures were, for pathological worry, the Penn State Worry Questionnaire (PSWQ) and, for anxiety and impairment, the Overall Anxiety Severity and Impairment Scale (OASIS).

Patients completed the GAD-7 at the beginning of each of the first 11 sessions and at the end of the final 12th session. The 2 secondary measures were filled at the beginning and end of treatment. Each participant received an invitation for a follow-up measurement 3 months after completing the treatment.

GAD 7-Item Scale

The GAD-7 is a short, 7-item self-report questionnaire developed to measure GAD diagnostic symptom criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition [30]. Its internal consistency (Cronbach α) has generally ranged between .88 in the clinical population and .92 in the general population [36,37]. The test–retest reliability in a study was 0.83 [30]. In this study, Cronbach α was .725 before the treatment and .905 after the treatment.

Penn State Worry Questionnaire

The PSWQ is a 16-item questionnaire designed to measure pathological worry, a signature feature of GAD [38]. Internal consistency has ranged from .86 to .95 among various samples [38–40]. PSWQ sensitivity to change is good but less salient than that of GAD-7 [41]. Test–retest reliability has been inconsistent (0.67–0.93) [39,42], so a midrange score of 0.80 was our choice. In our study, the Cronbach α was .865 before the treatment and .931 after the treatment.

Overall Anxiety Severity and Impairment Scale

The OASIS is a 5-item scale developed as a brief transdiagnostic measure of anxiety severity and impairment [43]. Internal consistency has ranged from .84 to .93 in 3 large clinical samples [44–46]. OASIS's sensitivity to change has been established, but its diagnostic accuracy is low [45,47]. Test–retest reliability has been 0.73 in undergraduate samples and 0.82 in clinical samples [43,45] with a midrange value of 0.78 used. Cronbach

α in this study was .736 before the treatment and .888 after the treatment.

Adherence

We measured the number of completed sessions and the time of therapy, defined as the time between the first and last GAD-7 in-therapy measurements. Only patients who completed all 12 sessions were categorized as completers. The program introduced the main theoretical framework in the first 4 sessions, and the later sessions focused mostly on further implementation of basic principles or introduced secondary content. Hence, we subcategorized noncompleters into two groups a priori: early dropouts (those discontinued before the completion of the fourth session) and late dropouts.

Patient Characteristics

We collected patient demographics such as age, registered sex, and municipality class. The municipality class was either urban or nonurban, according to the official Finnish classification [48].

Statistical Analysis

Our data were a convenience sample of all consented patients who entered treatment after February 2016. Our desired power was 80% and type I error was 0.05 for the main tests and 0.01 for post hoc tests. When comparing 2 independent means, a requirement to detect a small effect size, a minimum of 0.2, would be achieved with a sample of 435 for main tests and 647 for post hoc tests, assuming a 45%/55% completer/noncompleter distribution. For mixed model parameters, we set significance at a conservative $P < .01$, while ensuring that Akaike and Bayesian Information Criteria [49,50] do not indicate worsening fit. The final mixed model was built in a backward stepwise manner.

The primary outcome analysis involved a linear mixed random model that allowed growth modeling to account for the changing pace of recovery at different time points during therapy, the use of all available data, and different intercept estimations for each individual. GAD-7 served as the dependent variable, and we modeled a growth curve using linear, quadratic, and cubic terms of the GAD-7 observation time. To control for the effect of *dose* of and time on therapy, we also tried a full model with the main effects and interactions of the GAD-7 observation session. We also input the main effects of gender, age, referral source, municipality class, and completion status, along with their linear combinations with time.

The ITT and noncompleter groups' effect sizes for GAD-7 were estimated with mixed model–estimated marginal means. This approach generally performs better than a pure last observation carried forward (LOCF) at handling dropout bias, especially in larger samples ($n > 400$) [51]. The method used to calculate effect sizes was that of Morris and DeShon [52] (equations 13 and 12). Calculations also used the baseline SD and observed pre–post correlation.

We calculated clinical change indexes using the LOCF values. Reliable change was defined as Reliable Change Index (RCI [53]), and recovery was defined as transitioning below the clinical cutoff on the primary outcome measure, GAD-7. The

RCI and clinical cutoff together provide stringent criteria for both response and recovery [54]. Test–retest reliability obtained from earlier normative studies allowed the calculation of the RCI. We explored recovery for those with a baseline GAD-7 score of ≥ 10 to avoid confounding recovery rates with those of patients who had already fulfilled the criterion at baseline. In reliable recovery, the patient fulfilled the criteria for both clinical recovery and RCI. Full symptomatic recovery (remission) was defined as a posttreatment score < 5 on the GAD-7.

We used SPSS Statistics 22 (IBM Corporation) for the analyses [55].

Results

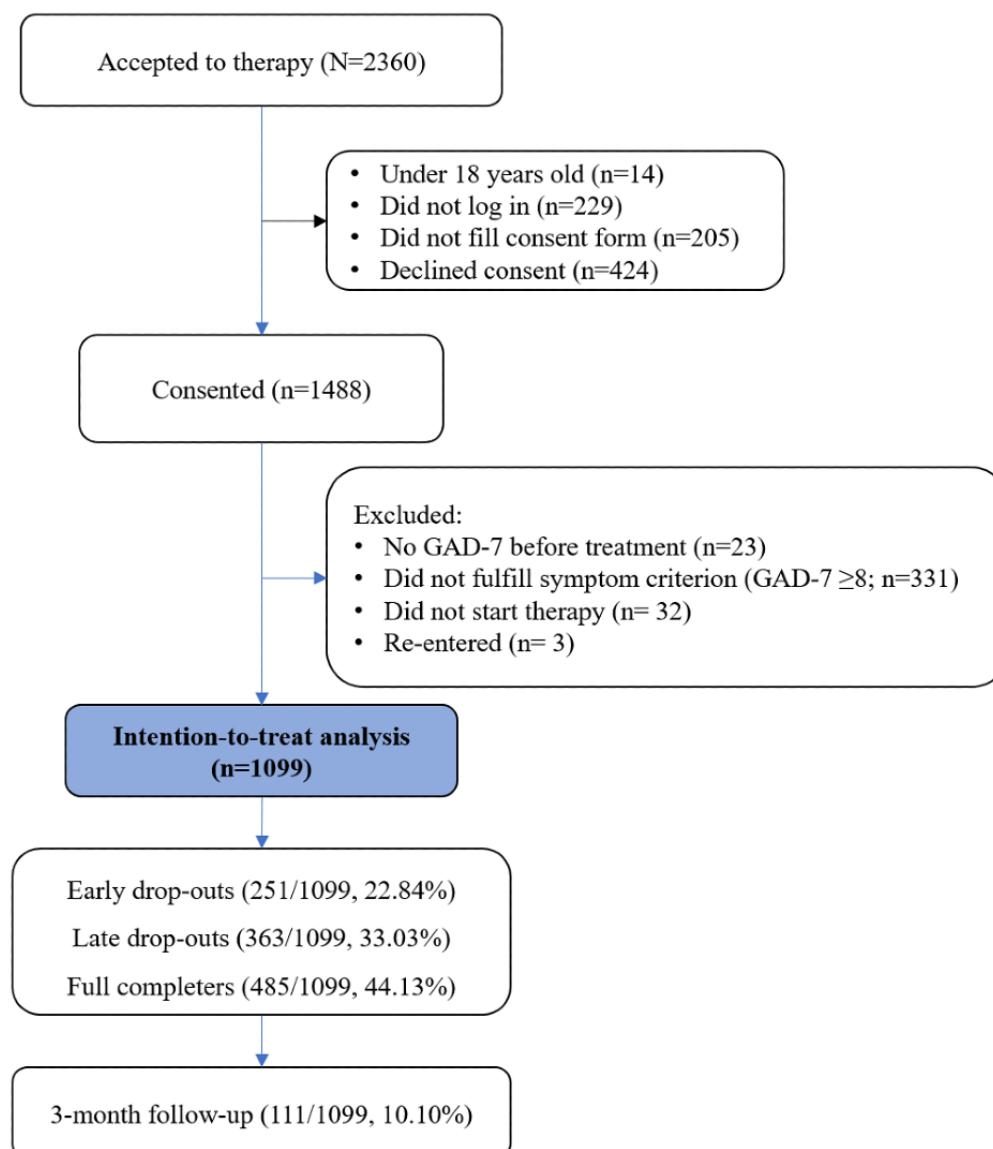
Patient Flow, Baseline Characteristics, and Adherence

Of all referrals to the clinic, those rejected during the recruitment period amounted to only 0.6%. Of the 1912 patients who completed the consent form, 1488 (77.82%) provided their consent (Figure 1).

Of the 1099 patients analyzed, 485 (44.13%) fully completed the HUS-iCBT, and a further 363 (33.03%) completed at least the first 4 sessions, meaning that 251 (22.84%) patients dropped out early. Those who completed from 4 to 11 sessions were considered as late dropouts. On average, patients completed 7.8 (SD 4.2; 65.1%) of the 12 sessions. The average time on therapy (time between pretreatment and last measurements) in the whole sample was 128 (SD 97) days, 171 (SD 95) days for completers, 36 (SD 44) days for early dropouts, and 131 (SD 84) days for late dropouts.

At baseline, completers and noncompleters differed in their proportions regarding referral source, municipality class, average age, and OASIS scores (Table 1). The average age of completers was 4.9 years higher with their average OASIS score being 0.7 points higher than the score of noncompleters. Statistically significant differences did not emerge in gender distribution or in the average GAD-7 and PSWQ scores.

Figure 1. Patient flowchart. GAD-7: Generalized Anxiety Disorder 7-item scale.



In post hoc analyses, completion was more likely among patients referred from private or occupational care (50.4%) than among those from primary care (39.1%; $\chi^2_1=9.675$; $P=.004$). Baseline OASIS scores differed significantly between early dropouts and both late dropouts ($t_{612}=4.297$; $P<.001$) and completers ($t_{734}=6.041$; $P<.001$), but not between late dropouts and completers ($t_{846}=1.625$; $P=.105$). The average baseline OASIS scores were 13.3 for early dropouts, 12.3 for late dropouts, and 12.0 for completers.

In total, 30 therapists treated an average of 37 (SD 38) study patients. Of the 1099 patients, 1097 (99.81%) received at least one message from their therapist, and 796 (72.43%) sent one or more messages to their therapist. Completers sent a message more often (420/485, 86.6% patients) than noncompleters (376/614, 61.2% patients). Therapists sent, on average, 8.4 (SD 4.7) messages and patients (those who did) sent 4.0 (SD 4.2) messages. Of those patients who did send messages, the average number of messages that completers sent (4.9, SD 4.8) was greater than that of noncompleters (3.1, SD 3.1). Therapists sent, on average, more messages to completers (11.6, SD 4.4) than to noncompleters (5.9, SD 3.1).

Table 1. Baseline characteristics.

Characteristics	Total sample (N=1099)	Completers (n=485)	Noncompleters (n=614)	Completers vs noncompleters	
				Test statistic	P value
Gender, n (%)				$\chi^2_1=2.240$.13
Female	849 (77.3)	385 (79.4)	464 (75.6)		
Male	250 (22.7)	100 (20.6)	150 (24.4)		
Referral source, n (%)				$\chi^2_4=16.872$.002
Primary care	601 (54.7)	235 (48.5)	366 (59.6)		
Private or occupational	339 (30.8)	171 (35.3)	168 (27.4)		
Student health care	50 (4.5)	24 (4.9)	26 (4.2)		
Psychiatry	55 (5)	25 (5.2)	30 (4.9)		
Unspecified	54 (4.9)	33 (6.8)	21 (3.4)		
Municipality class, n (%)				$\chi^2_1=8.469$.004
Urban	966 (87.9)	412 (84.9)	554 (90.2)		
Nonurban	130 (11.8)	73 (7.2)	57 (5.7)		
Age (years), mean (SD)	33.3 (12.2)	36.0 (13.1)	31.1 (12.2)	$t_{935}=6.578$	<.001
GAD-7 ^a , mean (SD)	13.2 (3.6)	13.1 (3.6)	13.3 (3.6)	$t_{1038}=0.941$.35
PSWQ ^b , mean (SD)	64.4 (9.8)	64.7 (9.9)	64.2 (9.7)	$t_{1031}=0.723$.47
OASIS ^c , mean (SD)	12.4 (2.8)	12.0 (2.8)	12.7 (2.8)	$t_{1046}=4.216$	<.001

^aGAD-7: Generalized Anxiety Disorder 7-item scale.

^bPSWQ: Penn State Worry Questionnaire.

^cOASIS: Overall Anxiety Severity and Impairment Scale.

Primary Outcome Model

The building of the primary mixed model is described in [Multimedia Appendix 1](#) and [Textbox 1](#). The model was built backward stepwise, with all terms of interest entered into the model. Eliminated terms were gender, both in average effects and time interaction; municipality class, both in average effects and time interaction; age in average effects; and completion status, both in average effects and session interaction.

The estimated fixed and random effects on GAD-7 are given in [Table 2](#), and the model-estimated mean trajectory is shown in [Figure 2](#). According to the model, after the average time between the pre- and posttreatment observations for completers (171, SD 95 days), the main effects of time on GAD-7 would amount to an average of 3.1 (95% CI 4.4-2.7) points of

improvement. The main average effect of the session was 0.523 (95% CI 0.592-0.455) points of improvement per session, or 6.7 (95% CI 5.5-7.1) points after all 12 sessions. Time had a negative interaction with session, and over a complete therapy (12 sessions), an extra week of therapy would amount to an average of 0.20 (95% CI 0.16-0.24) points of deterioration. Older patients improved more than their younger counterparts; for example, 40-year-old patients improved on average 0.3 points more (95% CI 0.1-0.5) than their 20-year-old counterparts after an average course of therapy. Patients from private or occupational health care had similar overall symptoms as patients from student health care (95% CI -0.6 to 1.6) but milder than those from specialized psychiatric care (95% CI 1.0-3.1) and primary health care (95% CI 0.4-1.4). Patients from private and occupational health care benefited similarly as patients from

psychiatric services (95% CI -0.12 to 0.17) but more than patients from student (95% CI 0.13 - 0.46) or primary health care (95% CI 0.05 - 0.19) per month on therapy. Completers and early dropouts benefited from the treatment similarly (95% CI -0.16

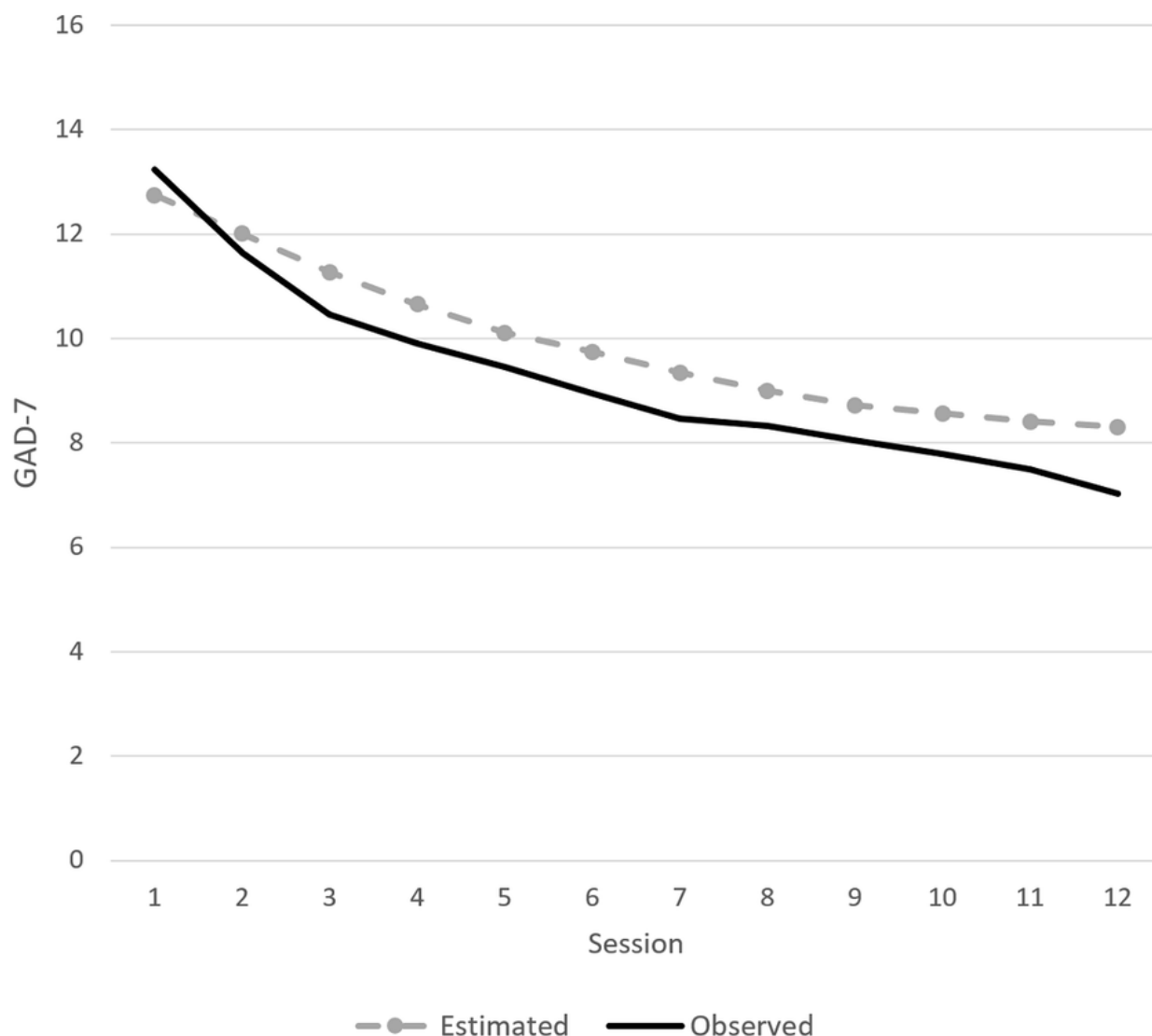
to 0.35), but late dropouts benefited less (95% CI 0.16 - 0.35) per month. Estimates of random effects indicated that patients likely had individual overall symptom levels ($P < .001$).

Table 2. Mixed linear model parameter estimates.

Parameter	Estimate (SE; 95% CI)	Wald Z	<i>t</i> test (<i>df</i>)	<i>P</i> value
Fixed effects		N/A ^a		
Intercept	12.422 (0.210; 12.01 to 12.83)		59.16 (1594.5)	<.001
Observation time (time)	-0.028 (0.004; -0.035 to -0.021)		-7.89 (8627.5)	<.001
Time ²	3.35×10^{-5} (1.32×10^{-5} ; 7.57×10^{-6} to 5.95×10^{-5})		2.53 (8308.9)	.01
Time ³	-3.39×10^{-8} (1.18×10^{-8} ; -5.70×10^{-8} to -1.07×10^{-8})		-2.87 (8117.6)	.004
Observation session (session)	-0.523 (0.035; -0.592 to -0.455)		-15.00 (8652.5)	<.001
Time \times session	0.002 (0.0003; 0.002 to 0.003)		9.11 (8240.2)	<.001
Time \times age	-1.32×10^{-4} (4.10×10^{-5} ; 2×10^{-4} to -5.11×10^{-5})		-3.21 (8669.3)	.001
Referral source, main effect		N/A		
Private or occupational	N/A		N/A	N/A
Psychiatry	2.059 (0.530; 1.018 to 3.099)		3.88 (1289.4)	<.001
Other	1.409 (0.533; 0.363 to 2.454)		2.64 (1272.5)	.008
Student health care	0.500 (0.559; -0.597 to 1.598)		0.89 (1336.0)	.37
Primary care	0.919 (0.249; 0.430 to 1.408)		3.68 (1303.0)	<.001
Referral source \times time		N/A		
Private or occupational	N/A		N/A	N/A
Psychiatry	0.001 (0.002; -0.004 to 0.006)		0.34 (8163.2)	.73
Other	0.004 (0.002; -0.001 to 0.009)		1.58 (8140.2)	.11
Student health care	0.010 (0.003; 0.004 to 0.015)		3.46 (8193.8)	.001
Primary care	0.004 (0.001; 0.002 to 0.006)		3.51 (8219.9)	<.001
Completion status \times time		N/A		
Completers, as reference	N/A		N/A	N/A
Late dropouts	0.005 (0.001; 0.002 to 0.008)		3.44 (8677.2)	.001
Early dropouts	0.003 (0.004; -0.005 to 0.012)		0.72 (8525.7)	.47
Random effects				
Residual	9.40 (0.15; 9.11 to 9.70)	61.76	N/A	<.001
Patient intercept	10.56 (0.52; 9.59 to 11.63)	20.34	N/A	<.001

^aN/A: not applicable.

Figure 2. Estimated mean marginal values based on linear mixed model versus observed values. Curves depict an average patient trajectory. GAD-7: Generalized Anxiety Disorder 7-item scale.



Effect Sizes

Treatment effect sizes are given in Table 3. The estimated ITT treatment effect size was 0.97, indicating a large effect on the GAD-7. The estimated scores indicated a small effect size in early dropouts (Cohen $d=0.34$) and a large effect size in late

dropouts (Cohen $d=0.85$). The treatment effect observed for the completers was very large (Cohen $d=1.39$).

At the 3-month follow-up, 111 completers were reached, which accounted for 10.1% of the whole sample and 22.9% of the completers. The observed change in the GAD-7 score was -6.8 . For secondary measures, changes were -11.3 in PSWQ and -3.9 in OASIS.

Table 3. Treatment effects based on estimated marginal means and observed values.

Treatment effect	Baseline, mean (SD)	Post, mean (SD)	Pre–post, change	Pre–post, correlation	Effect size, Cohen <i>d</i> (95% CI)
GAD-7^a					
Completers ^b	13.1 (3.6)	7.0 (4.7)	–6.1	0.267	1.39 (1.25-1.53)
Estimated values					
Early dropouts	12.9 (3.6)	11.5 (4.5)	–1.3	0.377 ^c	0.34 (0.16-0.51)
Late dropouts	12.9 (3.6)	9.4 (5.1)	–3.5	0.351 ^c	0.85 (0.70-1.00)
ITT ^d analysis	12.9 (3.6)	8.7 (4.7)	–4.1	0.308 ^c	0.97 (0.88-1.06)
PSWQ^e					
Completers ^b	64.7 (9.9)	53.7 (13.6)	–11.0	0.526	1.14 (1.00-1.27)
OASIS^f					
Completers ^b	12.0 (2.8)	8.4 (3.9)	–3.6	0.452	1.23 (1.09-1.37)

^aGAD-7: Generalized Anxiety Disorder 7-item scale.

^bFewer patients completed the PSWQ and OASIS after the treatment (n=479).

^cOnly these correlations were calculated from before the treatment and the last observed values for those with ≥2 observations.

^dITT: intention to treat.

^ePSWQ: Penn State Worry Questionnaire. For PSWQ and Overall Anxiety Severity and Impairment Scale, only baseline scores were available for noncompleters; hence, only complete effect sizes were observed.

^fOASIS: Overall Anxiety Severity and Impairment Scale.

Clinical Change Indexes

Clinical change indexes based on the GAD-7 scores are provided in Table 4. After the treatment, 23.29% (256/1099) of patients achieved remission (GAD-7 score <5).

Indexes for PSWQ and OASIS were reported only for completers because noncompleters had only one observation. Reliable change on PSWQ was a change of >12 points and 40.9% (196/479) reliably improved and 1% (5/479) reliably deteriorated. On OASIS reliable change was a change of >3 points and 48.6% (233/479) reliably improved and 1.5% (7/479) reliably deteriorated.

Table 4. Clinical indexes at the final on-therapy observation.

Clinical index	ITT ^a , n (%)	Completers, n (%)	Late dropouts ^b (sessions ≥4), n (%)	Early dropouts ^b (sessions <4), n (%)
Baseline GAD^c -7 score ≥8	1099 (100)	485 (100)	363 (100)	251 (100)
Reliable improvement ^d	532 (48.4)	306 (63.1)	170 (46.8)	56 (22.3)
Reliable deterioration ^e	41 (3.73)	12 (2.5)	19 (5.2)	10 (4)
Baseline GAD-7 score ≥10	894 (100)	385 (100)	295 (100)	214 (100)
Reliable improvement ^d	479 (53.6)	268 (69.6)	155 (52.5)	56 (26.2)
Recovery ^f	486 (54.4)	280 (72.7)	149 (50.5)	57 (26.6)
Reliable recovery ^g	421 (47.1)	244 (63.4)	129 (43.7)	48 (22.4)
Reliable deterioration ^e	27 (3)	9 (2.3)	11 (3.7)	7 (3.3)

^aITT: intention to treat.

^bOn the basis of the last observed values.

^cGAD-7: Generalized Anxiety Disorder 7-item scale.

^dGAD-7 score drop of ≥5.

^eGAD-7 score increase of ≥5.

^fGAD-7 changed to <10 at after the treatment.

^gBoth reliably improved and recovered.

Discussion

Principal Findings

This nationwide, free-of-charge, therapist-supported HUS-iCBT for GAD in routine care, with no predetermined maximum completion time, comprised 1099 patients referred by their physicians. To the best of our knowledge, this study has the largest reported real-world sample involving iCBT for GAD. The patients reported substantial improvement in ITT analyses amounting to effect sizes of Cohen $d=0.97$ on GAD-7; and for completers, Cohen $d=1.39$ on GAD-7, Cohen $d=1.14$ on PSWQ, and Cohen $d=1.23$ on OASIS. The full completion rate was 44.1% (485/1099), and the average session completion rate was 7.8 (65%) of the total 12 sessions. A reliable improvement occurred in 48.4% (532/1099) of the full sample. Of those with a baseline GAD-7 score ≥ 10 , reliable improvement occurred in 53.6% (479/894), with a transition to recovery achieved by 54.4% (421/894). These results seemed to be enduring for 22% (111/485) of completers who answered the follow-up questionnaires. The average change on the GAD-7 at 3-month follow-up was still -6.8 points (vs -6.1 at after the treatment).

Comparison With Earlier Work

Symptomatic Change

The change in GAD-7 score was comparable with that achieved in 3 of the 4 Australian ThisWayUp trials (Cohen $d=0.91-1.18$ [23,26,27]) lower than that in the fourth trial that used different inclusion criteria (Hedges $g=2.06$ and 2.10 [28]), and comparable with the Online Therapy Unit's trial (Cohen $d=1.07$ [29]). The change in GAD-7 scores in our study was also not inferior to the within-group change in a recent meta-analysis of RCTs of transdiagnostic iCBT for anxiety and depression (95% CI 0.91-1.22 [15]). The effect size also seems encouraging when contrasted with change in comparison with a waiting list in another meta-analysis (95% CI 0.39-1.01 [9]).

Interestingly, from both practical and theoretical viewpoints, the longer it took for patients to complete a given number of sessions, the less they improved. It is plausible that the effects of therapy become diluted if the patient's commitment to therapy weakens. Therefore, methods of increasing engagement in therapy seem one of the most promising ways to increase the effectiveness of iCBT for GAD.

The change in the PSWQ for completers (Cohen $d=1.14$) indicates a substantial improvement in worry. In a meta-analysis of 10 RCTs of iCBT for GAD, the ITT effect size on the PSWQ in iCBT was 0.71 [16]. Although the comparison is not direct, the results of this study agree with the existing RCT results.

The OASIS-measured anxiety-related functional impairment demonstrated a large improvement in the completers (Cohen $d=1.23$). Of the ThisWayUp trials, Mahoney et al [28] found medium to large improvements in health-related disability and functioning, as measured by the World Health Organization Disability Assessment Schedule 2.0. Similarly, a trial conducted by the Online Therapy Unit [29] found that patients had a large decrease in functional impairment after therapy, as measured by the Work and Social Adjustment Scale. The combined results

strongly suggest that iCBT for GAD can improve functional capacity.

Our completers improved faster than late dropouts, unlike those in 2 earlier studies, which showed no or negligible differences. However, significant improvement was also observed in our noncompleters, which could imply heterogeneity in this population. Feedback offered by the patients who withdrew from the Online Therapy Unit intervention also suggests that several patients may have discontinued treatment as their symptoms improved [29]. In iCBT for depression, there seems to be a subgroup of patients who benefit from the therapy rapidly and discontinue treatment prematurely [56]. These patients may regard further treatment as unnecessary. Hence, discontinuation should not be considered, by default, a failure.

Although not unequivocally proven, clinician-referred patients in iCBT tend to exhibit lower effect sizes than community-recruited or self-referred ones [13,57]. Thus, our results seem encouraging, as all patients in HUS-iCBT have been physician-referred.

Recovery and Reliable Change

In the ITT analysis using LOCF, of HUS-iCBT patients with a baseline GAD-7 score ≥ 10 , 54.4% (486/894) of patients achieved recovery. This recovery rate is comparable with that reported earlier in face-to-face CBT for GAD (51.4% [54]). Owing to differing imputation methods, we cannot directly compare our ITT results with those of Australian studies. Nevertheless, Hobbs et al [26] reported a 70% recovery rate for completers, which is comparable with our 72.7% (280/385).

Moreover, the criteria for reliable change in our study and previous studies differed, which prevented direct comparison. In one meta-analysis of RCTs of iCBT, reliable deterioration occurred in 3.1% (95% CI 1.5%-5.9%) of patients in trials for anxiety disorders [13]; in another meta-analysis of several disorders, reliable deterioration occurred in 5.8% of iCBT patients (vs 17.4% of patients on the waiting list) [21]. In HUS-iCBT, reliable deterioration occurred in 3.73% (41/1099) of patients. Thus, HUS-iCBT for GAD also appeared to be as safe as precious treatments and safer than no treatment.

Adherence

The full completion rate in our study (485/1099, 44.13%) was similar to the rates in previous routine care trials (36%-55%) but lower than that in a meta-analysis of routine care iCBT for depression and anxiety (61%) [13]. However, RCTs are likely to inflate adherence when compared with real-world trials. For instance, ThisWayUp's trials have shown a dramatic decrease from very high full completion rates when transitioning from a research setting into primary care (75%-85% vs 38% [24]).

The proportion of completed sessions in HUS-iCBT (65%) was marginally lower than that in the earlier trials (67%-77% in ThisWayUp's and 72% in Online Therapy Unit's trials) but somewhat higher than that in the previously mentioned meta-analysis (57% [13]). Compared with ThisWayUp's program, a lower proportion of completed sessions may be expected in a therapy comprising twice as many sessions (12 vs 6). Interestingly, more patients in the HUS-iCBT completed

≥6 sessions compared with the Australian trials (61% vs 36%-55%). On the other hand, the Online Therapy Unit had slightly better results with a program of similar length both in sessions ($n=12$) and average days on therapy (135 vs our 128). Differences in the course of therapy, recruitment, and therapist protocols prevent straightforward comparison. The comparison is further hampered by the overlapping recruiting periods in the Australian trials, as the number of patients reported remains unique. Hence, the optimal number of iCBT sessions for GAD remains unknown.

Compared with our noncompletion rate (614/1099, 55.86%), face-to-face individual psychotherapy for GAD has had a low dropout rate of 17% in a meta-analysis of RCT trials [58], and a meta-analysis of nonrandomized outpatient studies on CBT for adult anxiety disorders showed a similar dropout rate of 15% [59]. However, the criteria for completion in face-to-face psychotherapy are often more lenient, such as stopping when agreed upon with the therapist or completing a minimum number of sessions [59]. Thus, the adherence numbers are not comparable and are likely to be closer to iCBTs when analyzed with an equivalent metric.

In our study, completers sent and received more messages than did noncompleters. Our data do not allow exploration of causality, but this finding may be intuitively explained by the completer's longer time on therapy. In the Online Therapy Unit's trial [29], both patients and therapists sent more messages (on average 9.6 and 20.0, respectively) than ours (4.2 and 8.4, respectively). This difference is not surprising, given the differences in intake protocol (both clinician referral and community recruitment and intake interview vs strictly a physician's referral and only referral screening in HUS-iCBT) and therapist-contact intensity (weekly messages vs minimum 4 messages per therapy). These differences may also partly explain the higher adherence in the Online Therapy Unit's trial.

Patients' Background Effects

Older patients improved more and faster than younger patients. At first glance, this seems counterintuitive, as one could presume that young, digital native patients could feel more comfortable in the digital realm than the older ones. Nevertheless, 1 of the 3 ThisWayUp's trials also found that older age predicted larger improvement [23], whereas another found no such relationship [26], and the third trial found a relationship but left its direction unreported [27]. Furthermore, in a transdiagnostic study of iCBT for anxiety and depression, older age predicted greater improvement [60], whereas in a meta-analysis of various psychotherapies for GAD, older participants improved less than younger ones [61].

Moreover, younger patients were less likely to complete the program than their older counterparts. This relationship has been a common finding in iCBT for GAD in routine care [23,26,27] but not in face-to-face psychotherapy for GAD [58]. The average age in this study was 33 years, 6 or 7 years lower than that in the 3 ThisWayUp and single Online Therapy Unit study that reported their sample ages [23,26,27,29]. Overall, research on the age issue is scarce, and what still remains obscure is how age may or may not affect the results of different modes of therapy, such as face-to-face therapy or iCBT.

Gender had no significant influence on adherence in our study, similar to the findings of Australian studies. Systematic reviews of RCTs on face-to-face psychotherapy for GAD or CBT for anxiety disorders have also found no significant gender-adherence relationship [54,58].

The full completion rate was higher among patients referred from private or occupational health (171/339, 50.4%) than among those referred from primary care (235/601, 39.1%). Their symptoms improved faster than those referred from students or primary health care. These effects could be due to patient-group differences: patients referred from occupational health are, by definition, employed and presumably have a higher average level of functioning, higher socioeconomic status, and a lower likelihood of serious comorbidities. Moreover, in our recent study, physicians in occupational health displayed more interest in sending patients to iCBT than their primary care counterparts [62], which may transfer into patients' own expectations and motivation.

Patients from nonurban municipalities were more likely to complete the therapy (73/130, 56.2%) than those from urban municipalities (412/966, 42.7%). This relationship is somewhat opposed to the findings in the 2012 report by Mewton et al [23], where patients from urban settings were more likely to complete the therapy than were those from a rural setting. The causes of this difference are unknown and are likely to include multiple recently discovered factors, such as differences in technology adoption, financial concerns, and access to treatment [63]. Nevertheless, the municipality class did not demonstrate a significant impact on symptomatic improvement in the mixed model analysis.

Design Comparison

The results of this and earlier studies in routine care are, in general, comparable despite considerable differences in setting, design, and support.

To the best of our knowledge, HUS-iCBT and the Online Therapy Unit's programs had no predetermined maximum therapy completion time, whereas ThisWayUp applied a typical fixed maximum time restriction (90 days). The time from the first to last observation on HUS-iCBT (128 days) and from the first to last log-in at the Online Therapy Unit (135 days) were comparable and longer than the typical 90 days in other trials. Although the lack of a predetermined maximum time span may support adherence during changing life situations, it could also disengage some patients and dilute therapy effects, thereby leading to an increased dropout rate. As confounding design and sample features exist, such as different numbers of sessions and differences in sample average age, the optimal time span remains unknown.

In the Australian studies, patients were referred to iCBT by their own independent clinicians, each of whom was required to register as a provider, to receive training, and to use an assessment toolkit. The Online Therapy Unit required a centralized diagnostic interview. Our HUS-iCBT only screened referrals from physicians (but not from other clinicians) with no obligatory registration or any specific assessment schema. The practice at HUS-iCBT was chosen as the middle ground to

ensure proper diagnostics while maintaining a high intake flow. To maintain a low threshold, the inclusion criteria in HUS-iCBT were purposely loose, with only 0.6% (51/8394) of all physician referrals rejected during the recruitment period, and this may have contributed to a possible patient–therapy mismatch.

In HUS-iCBT, the support was provided centrally by specially trained and supervised mental health professionals, as recommended by authors from 5 universities who argue that iCBT requires specialized expertise and processes [64]. In the ThisWayUp studies, support was provided by the referring clinicians, who rarely contacted their patients. A pre-existing alliance could compensate for the clinicians' lack of specialized training.

HUS-iCBT was free of charge for the patients, as was the treatment at the Online Therapy Unit. Patients in ThisWayUp paid approximately Aus \$49 (US \$37) for the treatment, which may have ensured motivation and therefore improved adherence [24] but may also serve as a barrier to treatment.

Strengths and Limitations

The main strengths of this study are its nationwide scope, the largest sample thus far reported, and its routine care setting. One could argue that reliance on self-report measures can inflate treatment effects, but 2 recent meta-analyses suggest that combining indexes based on self-reporting can be even conservatively biased [7,54]. Self-report measures also fulfill practical requirements for scalability.

Other routine care studies did not use PSWQ or OASIS. Their addition in our study shows that iCBT for GAD can ease pathological worry and anxiety-related impairments in routine care. Although they both offer important information, they may not be as sensitive to change as the GAD-7.

Concomitant treatments, including psychopharmaceuticals, are often offered to patients in routine care. Information on these treatments is not reliable. This limits our understanding of the possible interactions between different treatments.

Not all comorbidities may have appeared in patient documents in referrals, and depressive symptoms were not measured. This issue is particularly relevant for GAD, as its annual comorbidity with major depression is around 41% [4]. One effectiveness study also indicated that depressive symptoms could decrease treatment effectiveness [26]. The lack of depression measurement does not weaken the study's results but highlights the question of whether every one of our patients received optimal treatment.

We could not, retrospectively, reliably identify the therapist behind each message sent to the patient. We know that due to vacation periods, illness, or leaving HUS, not all messages sent to a patient may have come from the same therapist. This prevented us from analyzing the effect of the individual therapist, their profession, or other relevant training.

Only 23% (111/485) of completers were reached for the 3-month follow-up. Such high attrition limits our conclusions regarding long-term effectiveness. However, we did not include follow-up data in the linear mixed model analyses, and these limited data do not compromise the main results.

Future Research

Thus far, therapist-supported iCBT for GAD has not been compared with other active treatment forms or no treatment in routine care, making comparisons with active controls, coupled with health economic analyses, essential. Not all patients benefit from iCBT for GAD. Both transdiagnostic programs and tailoring content to individual needs have displayed promising results [15,64], although for anxiety, unguided transdiagnostic iCBTs may be less or no more efficacious than diagnosis-specific ones [19,35]. Tailored transdiagnostic treatments seem likely to be a significant trend in iCBT, but evidence for their added value still needs to be confirmed.

Adherence to iCBT for GAD seems to be lower in primary care than in RCT trials. Evidence is emerging for adapted therapist support for patients at risk for dropout [65,66] and for an optimal therapist-contact schedule for iCBT in cases of depression and insomnia [67,68]. Machine learning applications could accelerate the identification of patients at risk for dropout and accelerate proactive support. However, this venue of research is still in its infancy.

iCBT programs often retain a considerable number of text-based answers and messages. Text mining approaches could offer an exciting avenue for exploring qualitative phenomena. Such analyses could be beneficial for both theoretical research and the practical application of iCBT.

The optimal time schedules for or the number of sessions in iCBT for GAD remain unclear. Further comparison studies should seek to establish more suitable, effective, and economic combinations. Age and other demographic variables may also influence the outcomes of iCBT for GAD. To provide equally effective care for everyone, future studies should investigate how to overcome hurdles presented by demographic variables.

PSWQ and OASIS were not used in other routine care studies. They represent both diagnosis-specific and transdiagnostic processes related to GAD and anxiety in general. OASIS, a short self-rating scale, opens up a venue for further comparative research between iCBT for various psychiatric disorders. As PSWQ measures pathological worry, a theoretically important GAD feature, an in-depth psychometric examination of PSWQ may reveal important information on moderators and mediators of effectiveness.

Since the collection of our data, HUS-iCBT for GAD has adopted a novel technical platform, a predetermined maximum completion time (20 weeks), pretreatment phone calls, and weekly therapist support. These changes may have practical implications that will enable further studies.

Conclusions

This nationwide, free-of-charge, therapist-supported HUS-iCBT for GAD with no predetermined maximum completion time was effective at improving symptoms and reducing worry and functional impairment in routine care. Overall, this therapy appears to be safer than no treatment and is at least as effective as other therapist-supported iCBTs for GAD. The observed gains at the 3-month follow-up should be confirmed in future studies. Future research needs to establish comparative

effectiveness against other treatments and to optimize the benefit of iCBT for GAD in a variety of patient groups and individual patients.

Acknowledgments

This work was supported by the Government of Finland and the Hospital Region of Helsinki and Uusimaa.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Description of primary mixed model building and statistics.

[[DOCX File, 15 KB - jmir_v24i3e29384_app1.docx](#)]

References

1. International Statistical Classification of Diseases and Related Health Problems, 10th Revision, Fifth Edition, 2016. Geneva: World Health Organization; 2015.
2. Munk-Jørgensen P, Allgulander C, Dahl AA, Foldager L, Holm M, Rasmussen I, et al. Prevalence of generalized anxiety disorder in general practice in Denmark, Finland, Norway, and Sweden. *Psychiatr Serv* 2006 Dec;57(12):1738-1744. [doi: [10.1176/ps.2006.57.12.1738](https://doi.org/10.1176/ps.2006.57.12.1738)] [Medline: [17158488](https://pubmed.ncbi.nlm.nih.gov/17158488/)]
3. Revicki DA, Travers K, Wyrwich KW, Svedäter H, Locklear J, Matterna MS, et al. Humanistic and economic burden of generalized anxiety disorder in North America and Europe. *J Affect Disord* 2012 Oct;140(2):103-112. [doi: [10.1016/j.jad.2011.11.014](https://doi.org/10.1016/j.jad.2011.11.014)] [Medline: [22154706](https://pubmed.ncbi.nlm.nih.gov/22154706/)]
4. Ruscio AM, Hallion LS, Lim CC, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, et al. Cross-sectional comparison of the epidemiology of DSM-5 generalized anxiety disorder across the globe. *JAMA Psychiatry* 2017 May 01;74(5):465-475 [FREE Full text] [doi: [10.1001/jamapsychiatry.2017.0056](https://doi.org/10.1001/jamapsychiatry.2017.0056)] [Medline: [28297020](https://pubmed.ncbi.nlm.nih.gov/28297020/)]
5. Generalised anxiety disorder and panic disorder in adults: management. National Institute for Health and Care Excellence. URL: <https://www.nice.org.uk/guidance/cg113/> [accessed 2022-02-13]
6. Anxiety Disorders. Current Care Guidelines. Working Group Set Up by the Finnish Medical Society Duodecim and the Finnish Psychiatric Association. Helsinki: The Finnish Medical Society Duodecim; 2019. URL: <https://www.kaypahoito.fi/hoi50119>
7. Cuijpers P, Sijbrandij M, Koole S, Huibers M, Berking M, Andersson G. Psychological treatment of generalized anxiety disorder: a meta-analysis. *Clin Psychol Rev* 2014 Mar;34(2):130-140. [doi: [10.1016/j.cpr.2014.01.002](https://doi.org/10.1016/j.cpr.2014.01.002)] [Medline: [24487344](https://pubmed.ncbi.nlm.nih.gov/24487344/)]
8. Titov N, Dear B, Niessen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. *Internet Interv* 2018 Sep;13:108-115 [FREE Full text] [doi: [10.1016/j.invent.2018.07.006](https://doi.org/10.1016/j.invent.2018.07.006)] [Medline: [30206525](https://pubmed.ncbi.nlm.nih.gov/30206525/)]
9. Andrews G, Basu A, Cuijpers P, Craske MG, McEvoy P, English CL, 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](https://doi.org/10.1016/j.janxdis.2018.01.001)] [Medline: [29422409](https://pubmed.ncbi.nlm.nih.gov/29422409/)]
10. Ophuis RH, Lokkerbol J, Heemskerk SC, van Balkom AJ, Hilgsmann M, Evers SM. Cost-effectiveness of interventions for treating anxiety disorders: a systematic review. *J Affect Disord* 2017 Mar 01;210:1-13. [doi: [10.1016/j.jad.2016.12.005](https://doi.org/10.1016/j.jad.2016.12.005)] [Medline: [27988373](https://pubmed.ncbi.nlm.nih.gov/27988373/)]
11. 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](https://doi.org/10.1080/16506073.2017.1401115)] [Medline: [29215315](https://pubmed.ncbi.nlm.nih.gov/29215315/)]
12. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FM, Furukawa TA, Cuijpers P, Individual Patient Data Meta-Analyses for Depression (IPDMA-DE) Collaboration, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry* 2021 Apr 01;78(4):361-371 [FREE Full text] [doi: [10.1001/jamapsychiatry.2020.4364](https://doi.org/10.1001/jamapsychiatry.2020.4364)] [Medline: [33471111](https://pubmed.ncbi.nlm.nih.gov/33471111/)]
13. Etzelmueller A, Vis C, Karyotaki E, Baumeister H, Titov N, Berking M, et al. Effects of internet-based cognitive behavioral therapy in routine care for adults in treatment for depression and anxiety: systematic review and meta-analysis. *J Med Internet Res* 2020 Aug 31;22(8):e18100 [FREE Full text] [doi: [10.2196/18100](https://doi.org/10.2196/18100)] [Medline: [32865497](https://pubmed.ncbi.nlm.nih.gov/32865497/)]
14. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. *JAMA Psychiatry* 2019 Jul 01;76(7):700-707 [FREE Full text] [doi: [10.1001/jamapsychiatry.2019.0268](https://doi.org/10.1001/jamapsychiatry.2019.0268)] [Medline: [30994877](https://pubmed.ncbi.nlm.nih.gov/30994877/)]

15. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. *Cogn Behav Ther* 2017 Jan;46(1):1-28. [doi: [10.1080/16506073.2016.1231219](https://doi.org/10.1080/16506073.2016.1231219)] [Medline: [27712544](https://pubmed.ncbi.nlm.nih.gov/27712544/)]
16. Richards D, Richardson T, Timulak L, McElvaney J. The efficacy of internet-delivered treatment for generalized anxiety disorder: a systematic review and meta-analysis. *Internet Interventions* 2015 Sep;2(3):272-282. [doi: [10.1016/j.invent.2015.07.003](https://doi.org/10.1016/j.invent.2015.07.003)]
17. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. *JAMA Psychiatry* 2017 Apr 01;74(4):351-359. [doi: [10.1001/jamapsychiatry.2017.0044](https://doi.org/10.1001/jamapsychiatry.2017.0044)] [Medline: [28241179](https://pubmed.ncbi.nlm.nih.gov/28241179/)]
18. 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](https://doi.org/10.1016/j.cpr.2012.02.004)] [Medline: [22466510](https://pubmed.ncbi.nlm.nih.gov/22466510/)]
19. Domhardt M, Geßlein H, von Rezori RE, Baumeister H. Internet- and mobile-based interventions for anxiety disorders: a meta-analytic review of intervention components. *Depress Anxiety* 2019 Mar;36(3):213-224. [doi: [10.1002/da.22860](https://doi.org/10.1002/da.22860)] [Medline: [30450811](https://pubmed.ncbi.nlm.nih.gov/30450811/)]
20. Pihlaja S, Stenberg J, Joutsenniemi K, Mehik H, Ritola V, Joffe G. Therapeutic alliance in guided internet therapy programs for depression and anxiety disorders - A systematic review. *Internet Interv* 2018 Mar;11:1-10 [FREE Full text] [doi: [10.1016/j.invent.2017.11.005](https://doi.org/10.1016/j.invent.2017.11.005)] [Medline: [30135754](https://pubmed.ncbi.nlm.nih.gov/30135754/)]
21. Rozental A, Magnusson K, Boettcher J, Andersson G, Carlbring P. For better or worse: an individual patient data meta-analysis of deterioration among participants receiving internet-based cognitive behavior therapy. *J Consult Clin Psychol* 2017 Feb;85(2):160-177. [doi: [10.1037/ccp0000158](https://doi.org/10.1037/ccp0000158)] [Medline: [27775414](https://pubmed.ncbi.nlm.nih.gov/27775414/)]
22. Hedman E, Ljótsson B, Kalso V, Hesser H, El Alaoui S, Kraepelin M, et al. Effectiveness of internet-based cognitive behaviour therapy for depression in routine psychiatric care. *J Affect Disord* 2014 Feb;155:49-58. [doi: [10.1016/j.jad.2013.10.023](https://doi.org/10.1016/j.jad.2013.10.023)] [Medline: [24238951](https://pubmed.ncbi.nlm.nih.gov/24238951/)]
23. Mewton L, Wong N, Andrews G. The effectiveness of internet cognitive behavioural therapy for generalized anxiety disorder in clinical practice. *Depress Anxiety* 2012 Oct;29(10):843-849. [doi: [10.1002/da.21995](https://doi.org/10.1002/da.21995)] [Medline: [22949296](https://pubmed.ncbi.nlm.nih.gov/22949296/)]
24. Hilvert-Bruce Z, Rossouw PJ, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. *Behav Res Ther* 2012 Aug;50(7-8):463-468. [doi: [10.1016/j.brat.2012.04.001](https://doi.org/10.1016/j.brat.2012.04.001)] [Medline: [22659155](https://pubmed.ncbi.nlm.nih.gov/22659155/)]
25. Framework for FDA's real-world evidence program. U.S. Food and Drug Administration. URL: <https://www.fda.gov/media/120060/download> [accessed 2022-02-13]
26. Hobbs MJ, Mahoney AE, Andrews G. Integrating iCBT for generalized anxiety disorder into routine clinical care: treatment effects across the adult lifespan. *J Anxiety Disord* 2017 Oct;51:47-54. [doi: [10.1016/j.janxdis.2017.09.003](https://doi.org/10.1016/j.janxdis.2017.09.003)] [Medline: [28926805](https://pubmed.ncbi.nlm.nih.gov/28926805/)]
27. Newby JM, Mewton L, Andrews G. Transdiagnostic versus disorder-specific internet-delivered cognitive behaviour therapy for anxiety and depression in primary care. *J Anxiety Disord* 2017 Mar;46:25-34. [doi: [10.1016/j.janxdis.2016.06.002](https://doi.org/10.1016/j.janxdis.2016.06.002)] [Medline: [27396841](https://pubmed.ncbi.nlm.nih.gov/27396841/)]
28. Mahoney AE, Newby JM, Hobbs MJ, Williams AD, Andrews G. Reducing behavioral avoidance with internet-delivered cognitive behavior therapy for generalized anxiety disorder. *Internet Interv* 2019 Mar;15:105-109 [FREE Full text] [doi: [10.1016/j.invent.2017.11.004](https://doi.org/10.1016/j.invent.2017.11.004)] [Medline: [30792961](https://pubmed.ncbi.nlm.nih.gov/30792961/)]
29. Hadjistavropoulos HD, Pugh NE, Nugent MM, Hesser H, Andersson G, Ivanov M, et al. Therapist-assisted internet-delivered cognitive behavior therapy for depression and anxiety: translating evidence into clinical practice. *J Anxiety Disord* 2014 Dec;28(8):884-893 [FREE Full text] [doi: [10.1016/j.janxdis.2014.09.018](https://doi.org/10.1016/j.janxdis.2014.09.018)] [Medline: [25445078](https://pubmed.ncbi.nlm.nih.gov/25445078/)]
30. 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](https://doi.org/10.1001/archinte.166.10.1092)] [Medline: [16717171](https://pubmed.ncbi.nlm.nih.gov/16717171/)]
31. Borkovec T. The nature, functions, and origins of worry. In: *Worrying: Perspectives on Theory, Assessment and Treatment*. Hoboken, New Jersey, United States: John Wiley & Sons; 1994.
32. Dugas MJ, Freeston MH, Ladouceur R. Intolerance of uncertainty and problem orientation in worry. *Cognit Ther Res* 1997;21(6):593-606. [doi: [10.1023/A:1021890322153](https://doi.org/10.1023/A:1021890322153)]
33. Wells A. The metacognitive model of worry and generalised anxiety disorder. In: *Worry and Its Psychological Disorders: Theory, Assessment and Treatment*. Hoboken, New Jersey, United States: Wiley; 2006.
34. Hayes S, Strosahl K, Wilson K. *Acceptance and Commitment Therapy, Second Edition The Process and Practice of Mindful Change*. New York, United States: Guilford Press; 2011.
35. McCall HC, Hadjistavropoulos HD, Sundström CR. Exploring the role of persuasive design in unguided internet-delivered cognitive behavioral therapy for depression and anxiety among adults: systematic review, meta-analysis, and meta-regression. *J Med Internet Res* 2021 Apr 29;23(4):e26939 [FREE Full text] [doi: [10.2196/26939](https://doi.org/10.2196/26939)] [Medline: [33913811](https://pubmed.ncbi.nlm.nih.gov/33913811/)]
36. Beard C, Björgvinsson T. Beyond generalized anxiety disorder: psychometric properties of the GAD-7 in a heterogeneous psychiatric sample. *J Anxiety Disord* 2014 Aug;28(6):547-552. [doi: [10.1016/j.janxdis.2014.06.002](https://doi.org/10.1016/j.janxdis.2014.06.002)] [Medline: [24983795](https://pubmed.ncbi.nlm.nih.gov/24983795/)]

37. Johansson R, Carlbring P, Heedman A, Paxling B, Andersson G. Depression, anxiety and their comorbidity in the Swedish general population: point prevalence and the effect on health-related quality of life. *PeerJ* 2013 Jul 9;1:e98 [FREE Full text] [doi: [10.7717/peerj.98](https://doi.org/10.7717/peerj.98)] [Medline: [23862109](https://pubmed.ncbi.nlm.nih.gov/23862109/)]
38. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. *Behav Res Ther* 1990;28(6):487-495. [doi: [10.1016/0005-7967\(90\)90135-6](https://doi.org/10.1016/0005-7967(90)90135-6)] [Medline: [2076086](https://pubmed.ncbi.nlm.nih.gov/2076086/)]
39. Zlomke KR. Psychometric properties of internet administered versions of Penn State Worry Questionnaire (PSWQ) and Depression, Anxiety, and Stress Scale (DASS). *Comput Human Behav* 2009 Jul;25(4):841-843. [doi: [10.1016/j.chb.2008.06.003](https://doi.org/10.1016/j.chb.2008.06.003)]
40. Verkuil B, Brosschot JF. The online version of the Dutch Penn State Worry Questionnaire: factor structure, predictive validity and reliability. *J Anxiety Disord* 2012 Dec;26(8):844-848. [doi: [10.1016/j.janxdis.2012.08.002](https://doi.org/10.1016/j.janxdis.2012.08.002)] [Medline: [23023163](https://pubmed.ncbi.nlm.nih.gov/23023163/)]
41. Dear BF, Titov N, Sunderland M, McMillan D, Anderson T, Lorian C, et al. Psychometric comparison of the generalized anxiety disorder scale-7 and the Penn State Worry Questionnaire for measuring response during treatment of generalised anxiety disorder. *Cogn Behav Ther* 2011;40(3):216-227. [doi: [10.1080/16506073.2011.582138](https://doi.org/10.1080/16506073.2011.582138)] [Medline: [21770844](https://pubmed.ncbi.nlm.nih.gov/21770844/)]
42. Molina S, Borkovec T. The Penn State Worry Questionnaire: Psychometric properties and associated characteristics. In: *Worrying: Perspectives on Theory, Assessment and Treatment*. Hoboken, New Jersey, United States: John Wiley & Sons; 1994.
43. Norman SB, Cissell SH, Means-Christensen AJ, Stein MB. Development and validation of an Overall Anxiety Severity And Impairment Scale (OASIS). *Depress Anxiety* 2006;23(4):245-249. [doi: [10.1002/da.20182](https://doi.org/10.1002/da.20182)] [Medline: [16688739](https://pubmed.ncbi.nlm.nih.gov/16688739/)]
44. Campbell-Sills L, Norman SB, Craske MG, Sullivan G, Lang AJ, Chavira DA, et al. Validation of a brief measure of anxiety-related severity and impairment: the Overall Anxiety Severity and Impairment Scale (OASIS). *J Affect Disord* 2009 Jan;112(1-3):92-101 [FREE Full text] [doi: [10.1016/j.jad.2008.03.014](https://doi.org/10.1016/j.jad.2008.03.014)] [Medline: [18486238](https://pubmed.ncbi.nlm.nih.gov/18486238/)]
45. Hermans M, Korrelboom K, Visser S. A Dutch version of the Overall Anxiety Severity and Impairment Scale (OASIS): psychometric properties and validation. *J Affect Disord* 2015 Feb 01;172:127-132. [doi: [10.1016/j.jad.2014.09.033](https://doi.org/10.1016/j.jad.2014.09.033)] [Medline: [25451406](https://pubmed.ncbi.nlm.nih.gov/25451406/)]
46. Ito M, Oe Y, Kato N, Nakajima S, Fujisato H, Miyamae M, et al. Validity and clinical interpretability of Overall Anxiety Severity And Impairment Scale (OASIS). *J Affect Disord* 2015 Jan 01;170:217-224. [doi: [10.1016/j.jad.2014.08.045](https://doi.org/10.1016/j.jad.2014.08.045)] [Medline: [25259673](https://pubmed.ncbi.nlm.nih.gov/25259673/)]
47. Moore SA, Welch SS, Michonski J, Poquiz J, Osborne TL, Sayrs J, et al. Psychometric evaluation of the Overall Anxiety Severity And Impairment Scale (OASIS) in individuals seeking outpatient specialty treatment for anxiety-related disorders. *J Affect Disord* 2015 Apr 01;175:463-470. [doi: [10.1016/j.jad.2015.01.041](https://doi.org/10.1016/j.jad.2015.01.041)] [Medline: [25679201](https://pubmed.ncbi.nlm.nih.gov/25679201/)]
48. Vuoden 2017 kuntien ja kuntaryhmistys-luokitukseen välinen luokitusavain. Tilastokeskus. URL: https://www2.tilastokeskus.fi/luokitukset/cormaps/kunta_1_20170101%23kuntaryhmitys_1_20170101/ [accessed 2022-02-13]
49. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Control* 1974 Dec;19(6):716-723. [doi: [10.1109/TAC.1974.1100705](https://doi.org/10.1109/TAC.1974.1100705)]
50. Schwarz G. Estimating the dimension of a model. *Ann Statist* 1978 Mar;6(2):461-464. [doi: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136)]
51. Mallinckrodt CH, Clark WS, David SR. Accounting for dropout bias using mixed-effects models. *J Biopharm Stat* 2001;11(1-2):9-21. [doi: [10.1081/BIP-100104194](https://doi.org/10.1081/BIP-100104194)] [Medline: [11459446](https://pubmed.ncbi.nlm.nih.gov/11459446/)]
52. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods* 2002 Mar;7(1):105-125. [doi: [10.1037/1082-989x.7.1.105](https://doi.org/10.1037/1082-989x.7.1.105)] [Medline: [11928886](https://pubmed.ncbi.nlm.nih.gov/11928886/)]
53. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. *J Consult Clin Psychol* 1991 Feb;59(1):12-19. [doi: [10.1037//0022-006x.59.1.12](https://doi.org/10.1037//0022-006x.59.1.12)] [Medline: [2002127](https://pubmed.ncbi.nlm.nih.gov/2002127/)]
54. Springer KS, Levy HC, Tolin DF. Remission in CBT for adult anxiety disorders: a meta-analysis. *Clin Psychol Rev* 2018 Apr;61:1-8. [doi: [10.1016/j.cpr.2018.03.002](https://doi.org/10.1016/j.cpr.2018.03.002)] [Medline: [29576326](https://pubmed.ncbi.nlm.nih.gov/29576326/)]
55. IBM Corp. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp; 2013.
56. Arndt A, Lutz W, Rubel J, Berger T, Meyer B, Schröder J, et al. Identifying change-dropout patterns during an internet-based intervention for depression by applying the Muthen-Roy model. *Cogn Behav Ther* 2020 Jan;49(1):22-40. [doi: [10.1080/16506073.2018.1556331](https://doi.org/10.1080/16506073.2018.1556331)] [Medline: [30721109](https://pubmed.ncbi.nlm.nih.gov/30721109/)]
57. 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](https://doi.org/10.2196/11706)] [Medline: [30994462](https://pubmed.ncbi.nlm.nih.gov/30994462/)]
58. Gersh E, Hallford DJ, Rice SM, Kazantzis N, Gersh H, Gersh B, et al. Systematic review and meta-analysis of dropout rates in individual psychotherapy for generalized anxiety disorder. *J Anxiety Disord* 2017 Dec;52:25-33. [doi: [10.1016/j.janxdis.2017.10.001](https://doi.org/10.1016/j.janxdis.2017.10.001)] [Medline: [29028610](https://pubmed.ncbi.nlm.nih.gov/29028610/)]
59. Hans E, Hiller W. A meta-analysis of nonrandomized effectiveness studies on outpatient cognitive behavioral therapy for adult anxiety disorders. *Clin Psychol Rev* 2013 Dec;33(8):954-964. [doi: [10.1016/j.cpr.2013.07.003](https://doi.org/10.1016/j.cpr.2013.07.003)] [Medline: [23988455](https://pubmed.ncbi.nlm.nih.gov/23988455/)]
60. Edmonds M, Hadjistavropoulos HD, Schneider LH, Dear BF, Titov N. Who benefits most from therapist-assisted internet-delivered cognitive behaviour therapy in clinical practice? Predictors of symptom change and dropout. *J Anxiety Disord* 2018 Mar;54:24-32 [FREE Full text] [doi: [10.1016/j.janxdis.2018.01.003](https://doi.org/10.1016/j.janxdis.2018.01.003)] [Medline: [29421369](https://pubmed.ncbi.nlm.nih.gov/29421369/)]

61. Carl E, Witcraft SM, Kauffman BY, Gillespie EM, Becker ES, Cuijpers P, et al. Psychological and pharmacological treatments for generalized anxiety disorder (GAD): a meta-analysis of randomized controlled trials. *Cogn Behav Ther* 2020 Jan;49(1):1-21 [FREE Full text] [doi: [10.1080/16506073.2018.1560358](https://doi.org/10.1080/16506073.2018.1560358)] [Medline: [30760112](https://pubmed.ncbi.nlm.nih.gov/30760112/)]
62. Taiminen HM, Saraniemi S, Joffe G, Stenberg J, Parkinson J. Reducing health inequalities through digital options in mental health: a physician's perspective. *Health Mark Q* 2019;36(2):93-106. [doi: [10.1080/07359683.2019.1575059](https://doi.org/10.1080/07359683.2019.1575059)] [Medline: [30907260](https://pubmed.ncbi.nlm.nih.gov/30907260/)]
63. Moskalenko MY, Hadjistavropoulos HD, Katapally TR. Barriers to patient interest in internet-based cognitive behavioral therapy: informing e-health policies through quantitative analysis. *Health Policy Technol* 2020 Jun;9(2):139-145. [doi: [10.1016/j.hlpt.2020.04.004](https://doi.org/10.1016/j.hlpt.2020.04.004)]
64. Titov N, Hadjistavropoulos HD, Niessen O, Mohr DC, Andersson G, Dear BF. From research to practice: ten lessons in delivering digital mental health services. *J Clin Med* 2019 Aug 17;8(8):1239 [FREE Full text] [doi: [10.3390/jcm8081239](https://doi.org/10.3390/jcm8081239)] [Medline: [31426460](https://pubmed.ncbi.nlm.nih.gov/31426460/)]
65. Forsell E, Jernelöv S, Blom K, Kraepelien M, Svanborg C, Andersson G, et al. Proof of concept for an adaptive treatment strategy to prevent failures in internet-delivered CBT: a single-blind randomized clinical trial with insomnia patients. *Am J Psychiatry* 2019 Apr 01;176(4):315-323. [doi: [10.1176/appi.ajp.2018.18060699](https://doi.org/10.1176/appi.ajp.2018.18060699)] [Medline: [30696270](https://pubmed.ncbi.nlm.nih.gov/30696270/)]
66. Pihlaja S, Lahti J, Lipsanen JO, Ritola V, Gummerus E, Stenberg J, et al. Scheduled telephone support for internet cognitive behavioral therapy for depression in patients at risk for dropout: pragmatic randomized controlled trial. *J Med Internet Res* 2020 Jul 23;22(7):e15732 [FREE Full text] [doi: [10.2196/15732](https://doi.org/10.2196/15732)] [Medline: [32706658](https://pubmed.ncbi.nlm.nih.gov/32706658/)]
67. Hadjistavropoulos HD, Schneider LH, Edmonds M, Karin E, Nugent MN, Dirkse D, et al. Randomized controlled trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. *J Anxiety Disord* 2017 Dec;52:15-24 [FREE Full text] [doi: [10.1016/j.janxdis.2017.09.006](https://doi.org/10.1016/j.janxdis.2017.09.006)] [Medline: [28964994](https://pubmed.ncbi.nlm.nih.gov/28964994/)]
68. Hadjistavropoulos HD, Schneider LH, Mehta S, Karin E, Dear BF, Titov N. Preference trial of internet-delivered cognitive behaviour therapy comparing standard weekly versus optional weekly therapist support. *J Anxiety Disord* 2019 Apr;63:51-60 [FREE Full text] [doi: [10.1016/j.janxdis.2019.02.002](https://doi.org/10.1016/j.janxdis.2019.02.002)] [Medline: [30844613](https://pubmed.ncbi.nlm.nih.gov/30844613/)]

Abbreviations

CBT: cognitive behavioral therapy
GAD: generalized anxiety disorder
GAD-7: Generalized Anxiety Disorder 7-item scale
iCBT: internet-delivered cognitive behavioral therapy
ITT: intention to treat
LOCF: last observation carried forward
OASIS: Overall Anxiety Severity and Impairment Scale
PSWQ: Penn State Worry Questionnaire
RCI: Reliable Change Index
RCT: randomized controlled trial

Edited by A Mavragani; submitted 05.04.21; peer-reviewed by H Hadjistavropoulos, S Victor, H Gandhi; comments to author 18.06.21; revised version received 30.09.21; accepted 16.12.21; published 24.03.22.

Please cite as:

Ritola V, Lipsanen JO, Pihlaja S, Gummerus EM, Stenberg JH, Saarni S, Joffe G
Internet-Delivered Cognitive Behavioral Therapy for Generalized Anxiety Disorder in Nationwide Routine Care: Effectiveness Study
J Med Internet Res 2022;24(3):e29384
URL: <https://www.jmir.org/2022/3/e29384>
doi: [10.2196/29384](https://doi.org/10.2196/29384)
PMID: [35323119](https://pubmed.ncbi.nlm.nih.gov/35323119/)

©Ville Ritola, Jari Olavi Lipsanen, Satu Pihlaja, Eero-Matti Gummerus, Jan-Henry Stenberg, Suoma Saarni, Grigori Joffe. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 24.03.2022. 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.

Original Paper

The Perceived Benefits of Digital Interventions for Behavioral Health: Qualitative Interview Study

Gabriela Marcu¹, PhD; Steven J Ondersma², PhD; Allison N Spiller¹, MSI; Brianna M Broderick¹, MSI; Reema Kadri³, MLIS; Lorraine R Buis^{1,3}, PhD

¹School of Information, University of Michigan, Ann Arbor, MI, United States

²Department of Obstetrics, Gynecology, & Reproductive Biology and the Division of Public Health, Michigan State University, East Lansing, MI, United States

³Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:

Lorraine R Buis, PhD

Department of Family Medicine

University of Michigan

1018 Fuller Street

Ann Arbor, MI, 48104-1213

United States

Phone: 1 734 998 7120

Fax: 1 734 998 7335

Email: buisl@umich.edu

Abstract

Background: Digital interventions have gained momentum in terms of behavioral health. However, owing to lacking standard approaches or tools for creating digital behavioral interventions, clinical researchers follow widely varying conceptions of how best to go about digital intervention development. Researchers also face significant cost-, time-, and expertise-related challenges in digital intervention development. Improving the availability of tools and guidance for researchers will require a thorough understanding of the motivations and needs of researchers seeking to create digital interventions.

Objective: This study aims to understand the perceptions of behavioral researchers toward digital interventions, and inform the use of these interventions, by documenting the reasons why researchers are increasingly focusing their efforts on digital interventions and their perspectives on the perceived benefits that digital approaches can provide for researchers and intervention recipients.

Methods: We conducted semistructured qualitative interviews with 18 researchers who had experience designing digital behavioral interventions or running studies with them. A convenience sample of interviewees was recruited from among users of the Computerized Intervention Authoring System platform, a web-based tool that facilitates the process of creating and deploying digital interventions in behavioral research. Interviews were conducted over teleconference between February and April 2020. Recordings from the interviews were transcribed and thematically analyzed by multiple coders.

Results: Interviews were completed with 18 individuals and lasted between 24 and 65 (mean 46.9, SD 11.3) minutes. Interviewees were predominantly female (17/18, 94%) and represented different job roles, ranging from researcher to project or study staff. Four major themes came out of the interviews concerning the benefits of digital interventions for behavioral health: convenience and flexibility for interventionists and recipients, support for implementing evidence-based interventions with fidelity, scaling and improving access to interventions, and getting a foot in the door despite stigma and disenfranchisement.

Conclusions: Interviewees described a number of important potential benefits of digital interventions, particularly with respect to scientific rigor, scalability, and overcoming barriers to reaching more people. There are complex considerations with regard to translating behavior change strategies into digital forms of delivery, and interventionists make individual, sometimes unexpected, choices with minimal evidence of their relative effectiveness. Future research should investigate how behavioral researchers can be supported in making these choices toward usability, ease of access, and approachability of digital interventions. Our study underscores the need for authoring platforms that can facilitate the process of creating and deploying digital interventions to reach their full potential for interventionists and recipients alike.

(*J Med Internet Res* 2022;24(3):e34300) doi:[10.2196/34300](https://doi.org/10.2196/34300)

KEYWORDS

computers; mobile apps; screening; brief interventions; diagnosis; computer-assisted/methods; surveys and questionnaires; motivational interviewing; therapy; computer-assisted/methods; implementation; qualitative; mobile phone

Introduction

Background

Digital behavioral interventions, which rely on digital technologies to promote behavior change and maintain health [1], have been growing rapidly over the past several decades. Digital approaches have unique potential benefits in promoting behavioral health, especially with regard to risky behaviors and sensitive or stigmatized topics (eg, drug and alcohol use [2-4], sexual health [5,6], and mood and anxiety [7]) that many are hesitant to discuss openly [8].

However, the development of such interventions requires significant expertise, planning, and financial resources [9-11]. Researchers interested in digital interventions may need more support in areas such as design guidance, software development, and financial or other logistical considerations. However, this support must be grounded in a deep understanding of the ways that investigators are using, or hope to use, digital interventions in their research. The purpose of this 2-part qualitative investigation is to better understand the motivations, needs, constraints, and experiences of researchers who have looked to digital interventions in their work addressing behavioral health. This study was conducted as part of a redesign of the Computerized Intervention Authoring System (CIAS) platform, a web-based authoring tool for researchers [12-14]. This study includes previous and current CIAS users and CIAS-naïve users with interests in developing interventions for substance use [3,4], mental health [15], and HIV [16]. This convenience sample enabled us to investigate the experiences of researchers who have used an authoring tool to assist with the development, deployment, and scaling of digital interventions.

Objectives

In this paper (part 1), we seek to document the reasons why these researchers used digital health interventions to address behavioral health. In a companion paper (part 2 [17]), we report common barriers to the intervention creators' experience and key considerations that go into the design and implementation of digital interventions.

Methods

Overview and Ethics Approval

We conducted semistructured interviews with former, current, and potential future CIAS users, including investigators and research study staff. All methods used in this study were approved by the University of Michigan human subjects review board (HUM00171197) and Wayne State University institutional review board (IRB-19-10-1340) and were reported following the COREQ (Consolidated Criteria for Reporting Qualitative Studies) guidelines.

CIAS Platform

CIAS is an internet-based intervention authoring tool that allows users to easily build, edit, and share web-based digital interventions without needing to have computer programming experience. The CIAS Health Insurance Portability and Accountability Act-compliant platform provides researchers with the ability to develop tailored and personalized text-based interventions that can be narrated by a variety of avatars, both human and animal, capable of multiple different actions and voices (supporting >40 different languages, with male and female versions of most, and a range of accents and dialects for some of the more commonly spoken languages). For researchers, CIAS supports screening functionality to determine study eligibility, in-platform randomization to multiple conditions, custom summary report generation, on-demand data access, SMS text messaging alerts flagging specified user inputs (eg, flags for suicidal ideation), instant language translation, a visual map showing intervention flow, and web analytics. CIAS supports intervention development features such as multiple question types, natural language reflections, branching and tailoring, and integration of video content. CIAS has been used by numerous research groups to develop a variety of digital interventions, many of which have focused on stigmatized behaviors such as substance use [15,18-21]. With support from the National Institutes of Health (EB028990), our team is redesigning CIAS from the ground up. This research was conducted as part of the redesign activities.

Participant Recruitment

Interviewees were recruited via email solicitations from a purposive convenience sample of CIAS users via snowball sampling and through email to investigators familiar with conducting digital interventions. Most participants had experience with motivational interviewing (MI), an evidence-based technique to support behavior change [22], and in interventions on a variety of behavioral health topics. To be eligible to participate in this study, participants were required to be aged at least 18 years.

Study Procedures and Data Collection

Owing to the COVID-19 pandemic, all interviews were conducted via teleconference and recorded for later transcription. Interviews were conducted in single sessions between February and April 2020 by 2 trained female study team members (BMB and ANS) with expertise in user experience design and methodology and with Master's of Science in Information degrees. The 2 interviewers took field notes during the interviews, which typically lasted no more than 60 minutes. Study participants did not know the interviewers and were not given prior access to interview questions or guides; however, all participants knew they were being interviewed as part of CIAS redesign efforts. At the beginning of each interview, the interviewers read a standard script focused on information pertaining to the study, and the interviewees gave verbal consent before the interview. At the conclusion of the interviews,

interviewees completed a brief electronic survey to assess their comfort with CIAS (if they had used it previously) and collect basic demographics. Participants were neither granted access to interview transcripts for review nor asked to comment on the findings. Interviewees were offered a US \$20 check, which was mailed to their home, for each interview. Interviews were conducted until saturation was reached.

Analysis

Analysis began with a debriefing between the 2 interviewers after each interview, during which they reviewed their individual notes and gradually synthesized data across interviews, looking for patterns. Interpretations of the data and identified patterns were discussed at meetings with the full study team every other week. This stage of analysis categorized participants within the common roles they played in their interactions with the CIAS or other digital interventions. Interview recordings were then transcribed, and inductive thematic analysis was conducted by 2 coders (BMB and ANS), including an initial round of full analysis that finalized a set of patterns that became the themes

of this paper, followed by a second round of analysis that then validated these themes and confirmed connections among them. Memoing was used throughout this process.

Results

Overview

We solicited 24 former and current CIAS users for participation in this study. Of the 24 participants, the interviews were conducted with 17 (71% response rate). In addition, we solicited additional CIAS-naive users familiar with digital interventions through email recruitment using the University of Michigan Department of Family Medicine listserv and recruited 1 additional participant. In total, we completed interviews with 18 participants. The interview duration lasted between 24 and 65 minutes (mean 46.9, SD 11.3 minutes). Interviewees were predominantly female (17/18, 94%) and represented different job roles, ranging from investigator to project or study staff. [Table 1](#) displays the aggregate interviewee characteristics, and [Table 2](#) displays brief descriptions of each interviewee.

Table 1. Aggregate interviewee characteristics (N=18).

Characteristics	Values, n (%)
Gender	
Male	1 (6)
Female	17 (94)
Job title (all that apply)	
Researcher	13 (72)
Psychologist	4 (22)
Other (project manager or coordinator and research assistant or associate)	5 (28)
Employer	
Academic institution	15 (83)
Foundation	1 (6)
Contract research organization	1 (6)
Other (nonprofit research organization)	1 (6)
Race	
African American or Black	4 (22)
White	13 (72)
Prefer not to answer	1 (6)
Ethnicity	
Hispanic or Latino	2 (11)
Non-Hispanic or Non-Latino	16 (89)
Education	
Bachelor's	6 (33)
Master's	4 (22)
Beyond a master's degree	8 (44)
Self-reported proficiency with CIAS^a (n=17)	
Novice	3 (18)
Proficient	4 (24)
Advanced	7 (41)
Expert	3 (18)

^aCIAS: Computerized Intervention Authoring System.

Table 2. Individual interviewee descriptions.

ID	Self-reported title	Employer type	Highest level of education obtained	Gender
P1	Researcher; psychologist	Academic institution	Beyond a master's degree	Female
P2	Researcher	Foundation	Master's degree	Female
P3	Project manager	Academic institution	Bachelor's degree	Female
P4	Psychologist	Academic institution	Beyond a master's degree	Female
P5	Project manager	Academic institution	Master's degree	Female
P6	Researcher	Academic institution	Beyond a master's degree	Male
P7	Researcher	Academic institution	Beyond a master's degree	Female
P8	Researcher; psychologist	Academic institution	Beyond a master's degree	Female
P9	Researcher; psychologist	Academic institution	Beyond a master's degree	Female
P10	Researcher	Academic institution	Bachelor's degree	Female
P11	Researcher	Academic institution	Bachelor's degree	Female
P12	Research assistant	Academic institution	Bachelor's degree	Female
P13	Researcher	Contract research organization	Master's degree	Female
P14	Researcher; research project coordinator	Academic institution	Master's degree	Female
P15	Researcher	Nonprofit research organization	Beyond a master's degree	Female
P16	Research associate	Academic institution	Bachelor's degree	Female
P17	Researcher	Academic institution	Bachelor's degree	Female
P18	Researcher	Academic institution	Beyond a master's degree	Female

Benefits of Digital Interventions

Overview

Our interviews revealed the perceived benefits of digital interventions from the perspective of researchers focusing on interventions for behavioral health. Grounding our interviews in the use of the CIAS platform led to themes focused on the concrete, current, or near-future capabilities of digital interventions. Interviewees most commonly pointed to four categories of potential: increased convenience and flexibility, implementing evidence-based interventions with fidelity, scaling interventions and improving access, and getting a foot in the door despite stigma and disenfranchisement.

Convenience and Flexibility for Interventionists and Recipients

Researchers described looking to digital interventions as they allow for more flexibility in where and when they are delivered. Interviewees discussed the potential for the digital approach to increase convenience and flexibility for researchers and the populations they are looking to reach. One of the researchers noted that the convenience of accessing a digital intervention is what an increasing number of generations is coming to expect of any service they receive:

We have water, we have food, and we have cell phones. Everyone is on their cell phones. Even the older generations have a phone. So, I think that CIAS is very convenient...I would like to see it a component like telemedicine is now. Telemedicine has its own phone number on the back of newer insurance cards

now...And I think that would definitely be rewarding to a lot of people. Because I see a lot of Millennials, and even the GenXers they're gravitating towards computer-based help. They're not jumping into their cars, spending \$3 for a gallon of gas to go to somebody's clinic. They're hopping online. [P10]

Evolving expectations about service delivery extended to the clinical point of view. Another researcher explained their view of how digital tools would likely support research and clinical visits in the future:

I think [CIAS] is going to allow us to deliver interventions at times that are more convenient for participants. As everything goes more mobile and more digital, being able to shoot a participant a link, enroll them in a study, shoot them a link, let them complete their follow up visits via that link, send me a text when they're done, shoot them an e-incentive. I think that's really going to change just delivering behavioral interventions in clinic and around clinic visits to delivering them, not when the client wants, so you're still on timed intervals, but when it's most convenient for the client. [P5]

However, researchers most often cited benefits for the recipients, including more choice in how they interact with the content. Recipients can complete intervention sessions wherever they are most comfortable and during a time they deem most appropriate while still staying within the allotted timeline of the intervention. This increase in flexibility may heighten the appeal of behavioral interventions to recipients, and researchers speculated that this could help with engagement:

I think a mobile intervention delivered on a computer is less threatening, and on the surface seems easier or more palatable to a research participant. "I don't have to go anywhere to do this, it can come to me at a time that's convenient to me"...The thought of coming into the office to do psychotherapy sessions is unappealing for many. But, looking at the convenience of my home at a time that is convenient for me, it may be more appealing. [P7]

Researchers were especially attuned to the needs of recipients who experienced symptoms such as fatigue, which could affect their engagement with an intervention. In these situations, interviewees had the sense that delivering interventions remotely via computers was helpful:

As far as it being remote, our participants, actually, the ones that do follow through and do their things, they love it because we're measuring their level of fatigue, so we aren't bothering them with making them come somewhere to have their visit completed. They can do it when they feel ready, feel able so it being remote is actually getting people to stay on, it's easier to get them to stay on because they don't have to go anywhere. [P11]

As we will discuss in following sections, the convenience of recipients accessing interventions from their homes is also important when addressing mental health and other sensitive or stigmatized topics.

Implementing Evidence-Based Interventions With Fidelity

Digital interventions also have the benefit of providing a structure that can support the use of evidence-based strategies with fidelity—that is, as intended or designed to ensure efficacy. Many of the researchers we interviewed were experienced in using MI and reflected on how and why they used digital interventions in concert with this approach. Digital interventions control the structure of how recipients experience the intervention, and researchers also remarked on how easily they could monitor the intervention and collect various kinds of data as recipients engage with it:

It's all programmed so there's less issues with fidelity and fidelity monitoring and so I guess that's the piece that changes CIAS, problem solving and/or developing strategies to monitor and track intervention fidelity. We still have to track delivery, so we still have to figure out "did they get it? did they complete it?" but we're not dealing with a human counselor who can kind of not follow the roadmap. [P7]

The analogy of an intervention as a road map suggests that systematic mapping of a digital intervention's user experience can serve to ossify the intervention and standardize how it is delivered to recipients. This mapping is performed once when the digital intervention is designed, and then researchers know that it will be delivered consistently each time. Therefore, interviewees indicated that improvements in fidelity could come

about not least by reducing reliance on humans, each of whom must be trained to effectively deliver the intervention:

Home visitors were mostly...they're not trained clinicians...they don't necessarily have the skills to really implement a brief motivational intervention in a way that it would be effective...so we sort of landed on CIAS where it's electronic and would be fully implemented electronically so the home visitor doesn't...have to learn how to do brief interventions. [P15]

Others similarly described how they had used digital interventions to overcome the variability inherent in training humans to deliver interventions, such as MI:

Most of the literature says that no matter how well you train people, that they don't deliver MI with high fidelity to the components that are supposed to be there at all. So, you spent all these hours and hours training people, and they still don't do it the way it's intended. So, of course, you know, having, and always doing it the same way makes a huge difference. [P1]

In addition to reliance on humans to deliver interventions consistently, researchers described challenges with training against the natural tendency to provide advice. By contrast, digital interventions were seen by interviewees as nonjudgmentally enabling a recipient to reflect on their health behaviors without the potential intrusion of another person's thoughts:

Our biggest challenge when teaching navigators...is getting them to not move into the role of "here's how to fix this and giving advice prematurely". And kind of having their own agenda. So, I think the pieces that can be done really well with CIAS are, you know, non-judgmental opportunities for a person to kind of look at their situation. [P9]

Interestingly, researchers saw an advantage in the fact that more nuanced aspects of MI, such as empathy, cannot be easily conveyed through a digital intervention. As recipients have lower expectations for warmth and understanding than they do of a human, a digital intervention may not need to meet many requirements for an effective interaction:

You can get away with a lot more on a computer-based intervention than you can with an in-person intervention. So, for instance, the idea of empathy and understanding, and warmth, and avoiding advice-giving and stuff like that, there are issues within person interaction. I don't think a lot of those are really problems with computer-based interventions. So, a computer can say "hey, wake up, go for a walk today" but if you told that to your roommate or boyfriend, that would be obnoxious, right? So, I wouldn't say that thinking is different about in-person than they are with online interventions, but the capacity of the two to kind of do the "dance" differently. [P6]

Other interviewees alluded to this dance between the intervention and the recipient, describing how they balance the

advantages and disadvantages of digital interventions to implement evidence-based strategies. For example, when using MI, there is a trade-off between the nuance a human is capable of and the difficulty in achieving consistent delivery of the intervention. Some researchers choose to sacrifice some of the former to improve the latter with a digital intervention:

The role of the avatar in this [intervention]...I think it isn't as good as a really good MI therapist delivering the intervention [using core principles like affirmations, empathy, non-judgement, no premature advice]. No, it's not, but as someone who trains people on a regular basis, you know the things that CIAS can do and do consistently, are things that I struggle to train caseworkers and patient navigators to be able to do. And I think in that sense it's a, it's a consistent step forward and an improvement...CIAS just lets you sort of balance that out [any variation in individual skill or technique] and make sure that everybody's getting some basic level of MI consistently. [P9]

Therefore, one of the key reasons that our interviewees chose to adopt digital interventions was to improve the fidelity of their evidence-based interventions through consistent delivery, monitoring, and avoidance of factors that are very difficult to manage when training humans. On the other hand, the downside of increasing structure and fidelity with digital interventions is losing the human ability to pick up on cues from the recipient, have more open-ended conversations, and adjust the intervention in response.

The Way of the Future: Scaling Interventions and Improving Access

Researchers were enthusiastic about the potential for scaling to broader populations with digital interventions in the future. This argument was typically framed from a population health perspective, as a relatively small change in behavior can have significant consequences if it can be achieved widely:

If you can reach more people and have a small change, that's important in shifting population behavior. So, digital mobile health interventions are the way of the future. [P7]

Similarly, researchers described the benefits to public health efforts if apps could bring interventions to those who are most difficult to reach:

I think it would be hugely impactful. I think a lot of the issues, or a lot of the barriers that public health faces as an industry is kind of getting people involved and actual...a lot of the issues with public health is low income and low resources for a lot of communities, and providing something so easy to use and so easy to get as an app, and allow public health professionals to communicate with so many people much easier and much, much more can make a world of difference. Program building and eliminating disease, and you know providing a more healthy lifestyle for people and kind of guiding them through that process. [P17]

One of the researchers further summarized the multilayered social factors that have resulted in populations being neglected and saw digital technologies as a way of addressing inequity:

There's this void within our community that a population is being ignored. A population that could find themselves in a mental health crisis. And they may be wheelchair bound. They may not have transportation. They may not have 24-7 care. So, a lot of the day-to-day operations of their life are online and on the phone. And it also fulfills a common item in our life. [P10]

The level of access a person has to services can also change significantly during a crisis in their personal life or their environment. In such situations, digital interventions can be more accessible and immediately available to provide tangible help if a person cannot leave their home during a crisis, as demonstrated by the global pandemic:

CIAS is catching up with the times and so to speak, providing interventions that are needed that are immediate. But in a convenient way. Like, for example, right now CIAS is needed because we're quarantined to our home [due to the COVID-19 pandemic]. Let's say for example, a person is on the verge of a break. They're in an emergency mental health crisis, they can't leave their home for whatever reason that may be. So, they can access the phone, get on the phone, and talk to somebody with real information with real resources. [P10]

With the ease of providing people access to interventions digitally, researchers highlighted the potential for scaling interventions as well as tailoring them to different people's needs:

[Digital interventions are] really scalable, you can tailor it in an infinite number of ways. [P6]

For example, recipients' literacy may be a consideration, in which case digital interventions could engage the recipients through audiovisual interaction:

I think illiteracy is an issue for some of these folks. So, I think having a survey they could participate in where they didn't have to worry about having to read things was something [important to us]. [P9]

The ability to reach people in new ways expanded researchers' thinking about new kinds of interventions. More than just translating existing interventions into digital formats, they were excited to think innovatively based on novel possibilities now afforded to them:

[CIAS] definitely broadened my ideas—like what is possible...like how many people could be reached by this. I think that especially—I am from rural Alabama and there's no way that you're ever going to hire like a—even a social worker is a stretch—but to get someone into a clinic setting who is trained and qualified to do these [assessments] is a stretch. And, you know, having an option to give someone a tablet while they're sitting there and they go through maybe a very quick screener and then they're eligible for

something. Or they're at risk for something and you're handing them something that's interactive and personalized and can give them feedback without having to have another person on the other end, I think is genius and could help a lot of people. [P16]

Some interviewees highlighted the potential for scaling access to interventions as the key reason they chose to use digital interventions:

And so, across all of those [potential benefits] for me the selling point has been getting interventions to more people more effectively. [P9]

However, we note that this potential remained largely aspirational for many researchers in our study, an indication of ongoing challenges with the development of digital interventions.

Getting a Foot in the Door Despite Stigma and Disenfranchisement

Access to health interventions is even more challenging when targeting behaviors that are stigmatized or populations who may not be comfortable seeking help. Disclosure of certain behaviors or circumstances is a significant barrier to connecting an individual with appropriate support. Several researchers mentioned that digital interventions enabled them to broach the taboo topic of substance use among pregnant and postpartum women. Therefore, by facilitating disclosure of sensitive issues such as substance use or domestic violence, digital interventions may open the possibility of intervening:

Being able to deliver an intervention online...you can send someone a link and they can go through a program on their computer, that could potentially really help them without having to disclose to a doctor or social worker or research assistant. I think that has massive implications for being able to get people help that's tailored to them. For substance use, I think it will help a lot...it has the potential in any kind of stigmatized, sensitive area, like in the doctor's office - a domestic violence screener. In any kind of healthcare setting trying to screen for any type of sensitive issue, I think it has a lot of potential to help in those areas.[P13]

Notably, disclosure was brought up in the context of both clinic and home visits. Researchers were cognizant of protecting confidentiality, whether engaging someone in a public clinical setting or via a visitor to their home. A digital intervention, even if handed over to the recipient by a human, was seen as potentially protective of their confidentiality:

We're hopeful that home visiting clients will be able to answer [sensitive] questions via an iPad, if you take away the idea that somebody else is judging you. [P2]

Researchers described intentionally deploying digital interventions to convey to recipients the confidentiality of any disclosures:

The client doesn't have to disclose if they don't want to, and that's a lot of the adapting that we're doing

is around that - making the client understand that this is confidential, that the home visitor doesn't need to know what they're putting into the program. [P15]

Barriers continue beyond screening and throughout the intervention process. Even if a potential intervention recipient can be identified (eg, through their disclosure) and linked to appropriate resources, they must be ready to accept information related to their health behavior. Therefore, the anonymity that digital interventions can afford may be helpful to an intervention itself getting a foot in the door, as one of the researchers explained within the context of engaging pregnant or postpartum individuals about their substance use:

These types of interventions, anonymous [and] in this population, do have the potential to get information and help to people who would not normally receive it because they would be too afraid to ask for it. I mean, just getting a foot in the door with these people and being able to give them psycho-education or resources at all, is a win and can really help people. [P13]

Interviewees also described using digital interventions with populations they would not have expected to reach in the past, for example, because of racial disparities or the stigma against seeking mental health services:

We're tapping into areas that we would have never thought of before. Like in terms of the disenfranchised populations [including based on their race], people who are afraid of being seen in the daylight entering into a mental health clinic. By having a phone, by having CIAS, we are getting them the help they need and we're giving them privacy...I do think we provide a comfort...I think our research becomes more and more validated because we're showing how flexible psychology and behavioral health can be. [P10]

Once an individual is engaged in an intervention, the relative neutrality of an app may enable them to dig deeper into a topic, less impeded by impression management in the presence of another human. One of the researchers working on substance use among new mothers was hopeful that switching to anonymous engagement with a digital intervention would help recipients feel more comfortable disclosing sensitive information about their behavior:

The most important thing is that CIAS allows us to have confidentiality and allows the mothers to go through the interventions anonymously. It has removed the barrier, because the intervention before...the home visitor had to go down and take the mother through it and they found that almost zero of the mothers were truthful because they didn't like talking to a person about this. So, having our little parrot narrator [avatar] bring them through it will hopefully lead to us getting accurate data for once. [P13]

As this researcher explained the ability of digital interventions to engage recipients differently, they also described the efficient progression possible by asking about certain health behaviors

to connect the recipient with appropriate resources in interactive and tailored ways:

In an in-person setting in this kind of thing, you can almost never get a straight answer if you ask somebody like “We saw that you just had a baby. Care to talk about your substance use while you were pregnant?” Like no, no one’s going to say anything about that, so you always have to come from a perspective of... “in the month before you became pregnant, what was your substance use like?” But in an anonymous format, we can kind of work around those proxy questions and get more to the meat of it, like “how are you feeling about your substance use?” “Do you want information on how you can change?” and then if so, it immediately pops up, like here’s your list of resources for your town. [A digital intervention] definitely changes the way that you can ask questions. [P13]

At the most basic level, interviewees acknowledged that intervention recipients might not even be receptive to the information provided by the intervention. In addition to removing the human interventionist and providing more privacy, digital interventions could address this issue by tailoring the information, how it is delivered, and when it is delivered:

I think one thing that CIAS has shown me is that people do appreciate the opportunity to learn information in private. And being that we thought they were absorbing or listening to what was being delivered from a person, they actually aren’t. Because when you validate the information that they should have got from a person, electronically you see that it doesn’t line up. I think CIAS has given me a new respect for being able to deliver information in a way people will receive it...since we’re using MI, one thing that’s pretty cool is that it does assess how much a person is willing to hear and it only gives them that and people tend to be interested in following up because they weren’t pressured into hearing something they didn’t want to hear and the time they didn’t want to hear it. [P5]

Some interviewees reported that digital interventions had shown marked improvement in the experiences of both researchers and intervention recipients:

We really have found CIAS to be preferable to the alternatives of both live person interview and certainly compared to survey questions without having the avatar. It just keeps people’s attention and tends to, in my experience, you know, continue to engage them in a way that they’re hopefully answering the questions more validly. [P9]

Such reports suggest that researchers have succeeded in getting their foot in the door with digital interventions, potentially even being better positioned to facilitate behavior change compared with other types of delivery. However, the use of digital interventions is still new, and research on their impact is limited.

Discussion

Principal Findings

We sought to better understand the reasons why some researchers who study topics related to behavioral health use digital interventions for lifestyle and behavior change interventions. The benefits of digital interventions that we identified among interviewees were 4-fold and included factors such as increased convenience and flexibility, fidelity to protocols, potential scalability and increased access to interventions, and the ability to reach individuals who may otherwise be unwilling to engage on matters of a sensitive nature.

Increased Convenience and Flexibility

Interviewees consistently noted that digital interventions increased convenience and flexibility for both researchers and intervention recipients. For researchers, digital interventions are set up in advance before study recruitment, and they typically run automatically without much staff involvement and oversight. For intervention recipients, depending on how digital interventions are delivered, there is potential to engage at a time and place that is convenient. Especially for sensitive issues, intervention recipients may not wish to engage in public settings and with research or clinic staff with whom they are unfamiliar.

Implementing Evidence-Based Interventions With Fidelity

Overview

The second theme that emerged from our interviews was the notion of fidelity to intervention protocols, both in terms of content and protocol fidelity. The algorithmic and automated nature of digital interventions can facilitate the delivery of content that is on schedule and consistently delivered in the same manner across intervention recipients. This protocol-based fidelity is more difficult with human interventionists.

Content Fidelity

Digital health interventions have the potential to ensure that content is delivered in the same way to all users, that is, in a nonjudgmental manner, which has been previously noted as a concern [23]. By digitally delivering content, we may be able to reduce social desirability bias. Moreover, automated interventions can finely tailor messages on a granular level in a way that is difficult for human interventionists.

Protocol Fidelity

The use of digital interventions can be tracked objectively and in detail. For example, log files and time stamps can verify when content is sent and when intervention recipients are engaged with the intervention. We can also objectively monitor measures of engagement, such as the number of times the content is accessed and the time spent engaging with content. Digital interventions allow information to be presented to recipients in a systematic way, which can allow for easier replication across various studies and with different populations, which could have a positive impact on the validity of a research study. This consistency can greatly benefit interventions that are based on

validated therapeutic techniques, such as MI, by standardizing its delivery across all recipients.

Fidelity Is Not Guaranteed

However, it should be noted that the benefits of content and protocol fidelity may not always be realized. Experienced researchers know that new threats to fidelity that are otherwise absent from face-to-face interventions exist in a digital world. Digital interventions often cannot guarantee that automated digital content is received by the correct person or that users engage with the content in the intended manner. Although technical logs can tell us that messages were sent and the amount of time spent within interfaces or on specific pages, researchers cannot guarantee that content was received or read, that the content reached the correct users, or that users attended to the content. Technical issues can also reduce the actual exposure to intervention content. Researchers new to digital interventions may not understand that routine and robust monitoring of the back end system status is an important requirement of digital interventions. Waiting on intervention recipients to report that an intervention is not properly working can lead to missed data collection and poor intervention fidelity. Being prepared to respond to system failures in a prompt fashion can reduce these potential issues. In their review of digital interventions adapting MI, Shingleton and Palfai [24] found that fidelity measures were often not reported in the literature.

Scaling Interventions and Improving Access

The third theme that emerged from our interviews was the unique potential of digital interventions to scale up and improve access to recipients. Regarding scalability, although initial costs are not insignificant for digital interventions, there is evidence that they can be cost-effective once developed [25]. However, robust data on digital interventions and cost-effectiveness are lacking [4,26-29]. From a research perspective, despite the high initial costs, digital interventions may reduce the need for costly human interventionists in research, which shifts human capital needs to technology development needs and may affect how money in research budgets is allocated. This increased potential for scalability means that population-based interventions are within better reach, and even small improvements in individual health outcomes can make big population-level improvements.

Issues related to access were especially important to the interviewees in this study. As noted earlier, digital interventions are potentially available at any time and any place. Moreover, as smartphone adoption is so ubiquitous, mobile interventions, or web-based interventions optimized for mobile devices, have the potential to reach even greater numbers than more traditional desktop-oriented programs. Although digital divide issues still persist as individuals with a lower socioeconomic status lag behind in terms of smartphone adoption, these populations with lower socioeconomic status tend to have better access to the internet via mobile phones than through computers and broadband in the home [30], which makes digital health interventions more important for reaching these populations. Digital interventions also have the potential to reach groups of people who may be hard to reach, such as those with physical disabilities, insufficient transportation, or significant time constraints. The nature of digital interventions may also increase

the potential to reach small, distributed groups of people who may have unique needs that could be too costly to support in person, such as individuals with rare diseases.

Beyond the potential to reach a greater number of people, digital interventions also have the potential to reach users in moments where they may be most efficacious [31]. For example, just-in-time adaptive interventions provide interventions at times where they may be most useful and at escalating doses that may lead to better outcomes [32]. Although most research in this area has been conducted with smoking cessation and physical activity and weight management interventions, just-in-time adaptive interventions have been previously used to intervene in sensitive behaviors such as risky substance use among adolescents and emerging adults [33]. Screening, Brief Intervention, and Referral to Treatment (SBIRT) interventions have also been used to intervene with intervention recipients at times when they may be available and receptive to interventions, such as before clinical encounters with primary care physicians or immediately following training sessions, and have been focused on topics such as substance use, intimate partner violence, and mental health [34]. These increases in convenience and flexibility for intervention recipients may yield downstream benefits to researchers and potentially translate to increased intervention engagement and completion, which may also yield positive improvements in health outcomes.

Getting a Foot in the Door Despite Stigma and Disenfranchisement

Findings from this study revealed that one perceived benefit of digital interventions for mental health and other stigmatized areas and behaviors is the opportunity to reach people concerning delicate topics when they are ready for and receptive to engaging in them. The anonymity provided by digital interventions may allow recipients to feel more comfortable disclosing sensitive information and engaging with related content. This may also afford users greater opportunities to delve into topics that they find more difficult to talk about. It is possible that this might lead to a greater level of honesty that could contribute to more successful intervention outcomes. Indeed, prior work has shown that self-reporting through digital formats can promote greater self-disclosure of undesirable behaviors and sensitive issues [8].

Future Directions

Given the range of potential advantages our interviewees described and the optimism and early gains they expressed, research is needed to understand the impact of delivering interventions digitally. For example, research should investigate the user experience from each stakeholder's point of view, identify potential unintended consequences, compare various digital and nondigital modalities for delivering interventions, and understand the disadvantages of digital delivery and how they can be addressed through careful intervention design and implementation. As mentioned in previous sections, research is needed to better understand the ways in which human-human interaction differs from human-computer interaction in terms of key factors such as perceived empathy, understanding, and friendliness.

Limitations

This study had several limitations that must be highlighted. This study was conducted as part of the redesign activities for CIAS, and therefore, it does not represent all researchers who use digital interventions for behavioral health. Moreover, our small sample size may have affected generalizability. As our convenience sample of researchers comprised CIAS users and those familiar with our redesign efforts, they may have felt compelled to aid our study activities, which could have introduced both selection bias and response bias. Finally, our focus on CIAS users meant that we only spoke to those individuals who were responsible for creating and managing digital interventions. Although many of our interviewees have conducted qualitative work with their own recipients to understand their perspectives, we did not speak directly with

recipients of digital interventions. Our findings are limited to the perceptions of digital intervention creators and study teams.

Conclusions

Digital interventions are seen by interventionists as having many benefits for both interventionists and recipients, such as greater flexibility in how, where, and when an intervention is delivered; potentially improving the fidelity of and access to interventions; and overcoming challenges of reaching populations despite stigma and disenfranchisement. More research is needed to measure the actual benefits of digital interventions and better understand whether they align with researcher expectations and hopes. A comparison of perceived and actual benefits may also inform better support for creating digital behavioral interventions that more effectively meet their potential.

Acknowledgments

This research was funded by the National Institute of Biomedical Imaging and Bioengineering at the National Institutes of Health under grant number EB028990 to SJO. The authors would like to thank Lilly Pritula and Rania Ajilat for administrative support, including manuscript editing, and the interviewees who made this study possible.

Conflicts of Interest

None declared.

References

1. Yardley L, Choudhury T, Patrick K, Michie S. Current issues and future directions for research into digital behavior change interventions. *Am J Prev Med* 2016 Nov;51(5):814-815. [doi: [10.1016/j.amepre.2016.07.019](https://doi.org/10.1016/j.amepre.2016.07.019)] [Medline: [27745680](https://pubmed.ncbi.nlm.nih.gov/27745680/)]
2. Fadus MC, Squeglia LM, Valadez EA, Tomko RL, Bryant BE, Gray KM. Adolescent substance use disorder treatment: an update on evidence-based strategies. *Curr Psychiatry Rep* 2019 Sep 14;21(10):96 [FREE Full text] [doi: [10.1007/s11920-019-1086-0](https://doi.org/10.1007/s11920-019-1086-0)] [Medline: [31522280](https://pubmed.ncbi.nlm.nih.gov/31522280/)]
3. Hankin A, Haley L, Baugher A, Colbert K, Houry D. Kiosk versus in-person screening for alcohol and drug use in the emergency department: patient preferences and disclosure. *West J Emerg Med* 2015 Mar 23;16(2):220-228 [FREE Full text] [doi: [10.5811/westjem.2015.1.24121](https://doi.org/10.5811/westjem.2015.1.24121)] [Medline: [25834660](https://pubmed.ncbi.nlm.nih.gov/25834660/)]
4. Kaner E, Beyer F, Garnett C. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. *Cochrane Database Syst Rev* 2017;9:CD011479. [doi: [10.1002/14651858.cd011479](https://doi.org/10.1002/14651858.cd011479)]
5. Kirana P, Gudeloglu A, Sansone A, Fode M, Reisman Y, Corona G, et al. E-Sexual health: a position statement of the European society for sexual medicine. *J Sex Med* 2020 Jul;17(7):1246-1253. [doi: [10.1016/j.jsxm.2020.03.009](https://doi.org/10.1016/j.jsxm.2020.03.009)] [Medline: [32340920](https://pubmed.ncbi.nlm.nih.gov/32340920/)]
6. Nadarzynski T, Morrison L, Bayley J, Llewellyn C. The role of digital interventions in sexual health. *Sex Transm Infect* 2017 Jun 08;93(4):234-235. [doi: [10.1136/sextrans-2016-052926](https://doi.org/10.1136/sextrans-2016-052926)] [Medline: [27932599](https://pubmed.ncbi.nlm.nih.gov/27932599/)]
7. Bucci S, Schwannauer M, Berry N. The digital revolution and its impact on mental health care. *Psychol Psychother* 2019 Jun 28;92(2):277-297. [doi: [10.1111/papt.12222](https://doi.org/10.1111/papt.12222)] [Medline: [30924316](https://pubmed.ncbi.nlm.nih.gov/30924316/)]
8. Gnambs T, Kaspar K. Disclosure of sensitive behaviors across self-administered survey modes: a meta-analysis. *Behav Res Methods* 2015 Dec 20;47(4):1237-1259. [doi: [10.3758/s13428-014-0533-4](https://doi.org/10.3758/s13428-014-0533-4)] [Medline: [25410404](https://pubmed.ncbi.nlm.nih.gov/25410404/)]
9. Cunha BC, Rodrigues KR, Zaine I, da Silva EA, Viel CC, Pimentel MD. Experience sampling and programmed intervention method and system for planning, authoring, and deploying mobile health interventions: design and case reports. *J Med Internet Res* 2021 Jul 12;23(7):e24278 [FREE Full text] [doi: [10.2196/24278](https://doi.org/10.2196/24278)] [Medline: [34255652](https://pubmed.ncbi.nlm.nih.gov/34255652/)]
10. Tate DF, Finkelstein EA, Khavjou O, Gustafson A. Cost effectiveness of internet interventions: review and recommendations. *Ann Behav Med* 2009 Aug 17;38(1):40-45 [FREE Full text] [doi: [10.1007/s12160-009-9131-6](https://doi.org/10.1007/s12160-009-9131-6)] [Medline: [19834778](https://pubmed.ncbi.nlm.nih.gov/19834778/)]
11. Biagianni B, Hidalgo-Mazzei D, Meyer N. Developing digital interventions for people living with serious mental illness: perspectives from three mHealth studies. *Evid Based Ment Health* 2017 Nov 12;20(4):98-101 [FREE Full text] [doi: [10.1136/eb-2017-102765](https://doi.org/10.1136/eb-2017-102765)] [Medline: [29025862](https://pubmed.ncbi.nlm.nih.gov/29025862/)]
12. Martino S, Ondersma SJ, Forray A, Olmstead TA, Gilstad-Hayden K, Howell HB, et al. A randomized controlled trial of screening and brief interventions for substance misuse in reproductive health. *Am J Obstet Gynecol* 2018 Mar;218(3):322.e1-322.e12 [FREE Full text] [doi: [10.1016/j.ajog.2017.12.005](https://doi.org/10.1016/j.ajog.2017.12.005)] [Medline: [29247636](https://pubmed.ncbi.nlm.nih.gov/29247636/)]

13. Ondersma SJ, Beatty JR, Svikis DS, Strickler RC, Tzilos GK, Chang G, et al. Computer-delivered screening and brief intervention for alcohol use in pregnancy: a pilot randomized trial. *Alcohol Clin Exp Res* 2015 Jul 22;39(7):1219-1226 [FREE Full text] [doi: [10.1111/acer.12747](https://doi.org/10.1111/acer.12747)] [Medline: [26010235](https://pubmed.ncbi.nlm.nih.gov/26010235/)]
14. Ondersma SJ, Svikis DS, Thacker LR, Beatty JR, Lockhart N. Computer-delivered screening and brief intervention (e-SBI) for postpartum drug use: a randomized trial. *J Subst Abuse Treat* 2014 Jan;46(1):52-59 [FREE Full text] [doi: [10.1016/j.jsat.2013.07.013](https://doi.org/10.1016/j.jsat.2013.07.013)] [Medline: [24051077](https://pubmed.ncbi.nlm.nih.gov/24051077/)]
15. Ahmedani BK, Crotty N, Abdulhak MM, Ondersma SJ. Pilot feasibility study of a brief, tailored mobile health intervention for depression among patients with chronic pain. *Behav Med* 2015 Sep 09;41(1):25-32. [doi: [10.1080/08964289.2013.867827](https://doi.org/10.1080/08964289.2013.867827)] [Medline: [24313728](https://pubmed.ncbi.nlm.nih.gov/24313728/)]
16. Naar-King S, Outlaw AY, Sarr M, Parsons JT, Belzer M, Macdonell K, Adolescent Medicine Network for HIV/AIDS Interventions. Motivational Enhancement System for Adherence (MESA): pilot randomized trial of a brief computer-delivered prevention intervention for youth initiating antiretroviral treatment. *J Pediatr Psychol* 2013 Jul 28;38(6):638-648 [FREE Full text] [doi: [10.1093/jpepsy/jss132](https://doi.org/10.1093/jpepsy/jss132)] [Medline: [23359664](https://pubmed.ncbi.nlm.nih.gov/23359664/)]
17. Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR. Barriers to and Considerations in the Design and Implementation of Digital Behavioral Interventions: Qualitative Analysis. *J Med Internet Res* 2022;24(3):e34301 [FREE Full text] [doi: [10.2196/34301](https://doi.org/10.2196/34301)]
18. Brown RA, Hecht J, Bloom EL, Minami H, Kahler CW, Abrantes AM, et al. Development and preliminary pilot evaluation of a brief tablet computer intervention to motivate tobacco quitline use among smokers in substance use treatment. *Am J Addict* 2017 Sep 11;26(6):587-594 [FREE Full text] [doi: [10.1111/ajad.12559](https://doi.org/10.1111/ajad.12559)] [Medline: [28800182](https://pubmed.ncbi.nlm.nih.gov/28800182/)]
19. Ellis DA, Carcone AI, Ondersma SJ, Naar-King S, Dekelbab B, Moltz K. Brief computer-delivered intervention to increase parental monitoring in families of African American adolescents with type 1 diabetes: a randomized controlled trial. *Telemed J E Health* 2017 Jun;23(6):493-502 [FREE Full text] [doi: [10.1089/tmj.2016.0182](https://doi.org/10.1089/tmj.2016.0182)] [Medline: [28061319](https://pubmed.ncbi.nlm.nih.gov/28061319/)]
20. Gryczynski J, Mitchell SG, Schwartz RP, Dusek K, O'Grady KE, Cowell AJ, et al. Computer- vs. nurse practitioner-delivered brief intervention for adolescent marijuana, alcohol, and sex risk behaviors in school-based health centers. *Drug Alcohol Depend* 2021 Jan 01;218:108423. [doi: [10.1016/j.drugalcdep.2020.108423](https://doi.org/10.1016/j.drugalcdep.2020.108423)] [Medline: [33307377](https://pubmed.ncbi.nlm.nih.gov/33307377/)]
21. Yonkers KA, Dailey JI, Gilstad-Hayden K, Ondersma SJ, Forray A, Olmstead TA, et al. Abstinence outcomes among women in reproductive health centers administered clinician or electronic brief interventions. *J Subst Abuse Treat* 2020 Jun;113:107995 [FREE Full text] [doi: [10.1016/j.jsat.2020.02.012](https://doi.org/10.1016/j.jsat.2020.02.012)] [Medline: [32359666](https://pubmed.ncbi.nlm.nih.gov/32359666/)]
22. Miller WR, Rollnick S. *Motivational Interviewing: Helping People Change*. 3rd ed. New York, NY: Guilford Press; 2012.
23. Aronson I, Freeman R, Taylor T, Bennett A. Developing digital media to destigmatize emergency department human immunodeficiency virus testing among sexual and racial minority youth: a hyper-iterative methodology. *Cureus* 2020 Mar 08;12(3):e7209 [FREE Full text] [doi: [10.7759/cureus.7209](https://doi.org/10.7759/cureus.7209)] [Medline: [32269887](https://pubmed.ncbi.nlm.nih.gov/32269887/)]
24. Shingleton RM, Palfai TP. Technology-delivered adaptations of motivational interviewing for health-related behaviors: a systematic review of the current research. *Patient Educ Couns* 2016 Jan;99(1):17-35 [FREE Full text] [doi: [10.1016/j.pec.2015.08.005](https://doi.org/10.1016/j.pec.2015.08.005)] [Medline: [26298219](https://pubmed.ncbi.nlm.nih.gov/26298219/)]
25. Jiang X, Ming W, You JH. The cost-effectiveness of digital health interventions on the management of cardiovascular diseases: systematic review. *J Med Internet Res* 2019 Jun 17;21(6):e13166 [FREE Full text] [doi: [10.2196/13166](https://doi.org/10.2196/13166)] [Medline: [31210136](https://pubmed.ncbi.nlm.nih.gov/31210136/)]
26. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital support interventions for the self-management of low back pain: a systematic review. *J Med Internet Res* 2017 May 21;19(5):e179 [FREE Full text] [doi: [10.2196/jmir.7290](https://doi.org/10.2196/jmir.7290)] [Medline: [28550009](https://pubmed.ncbi.nlm.nih.gov/28550009/)]
27. Rose T, Barker M, Maria Jacob C, Morrison L, Lawrence W, Strømmer 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](https://doi.org/10.1016/j.jadohealth.2017.05.024)] [Medline: [28822682](https://pubmed.ncbi.nlm.nih.gov/28822682/)]
28. Semwal M, Whiting P, Bajpai R, Bajpai S, Kyaw BM, Tudor Car L. Digital education for health professions on smoking cessation management: systematic review by the digital health education collaboration. *J Med Internet Res* 2019 Mar 04;21(3):e13000 [FREE Full text] [doi: [10.2196/13000](https://doi.org/10.2196/13000)] [Medline: [30829576](https://pubmed.ncbi.nlm.nih.gov/30829576/)]
29. Espinoza AV, Steurbaut S, Dupont A, Cornu P, van Hooff RJ, Brouns R, et al. Health economic evaluations of digital health interventions for secondary prevention in stroke patients: a systematic review. *Cerebrovasc Dis Extra* 2019 Jan 7;9(1):1-8 [FREE Full text] [doi: [10.1159/000496107](https://doi.org/10.1159/000496107)] [Medline: [30616238](https://pubmed.ncbi.nlm.nih.gov/30616238/)]
30. Vogels E. Digital divide persists even as Americans with lower incomes make gains in tech adoption. Pew Research Center. 2021. URL: <https://pewrsr.ch/2TRM7cP> [accessed 2021-10-12]
31. Linardon J, Messer M, Lee S, Rosato J. Perspectives of e-health interventions for treating and preventing eating disorders: descriptive study of perceived advantages and barriers, help-seeking intentions, and preferred functionality. *Eat Weight Disord* 2021 May 21;26(4):1097-1109. [doi: [10.1007/s40519-020-01005-3](https://doi.org/10.1007/s40519-020-01005-3)] [Medline: [32959274](https://pubmed.ncbi.nlm.nih.gov/32959274/)]
32. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: an experimental design for developing just-in-time adaptive interventions. *Health Psychol* 2015 Dec;34S(Suppl):1220-1228 [FREE Full text] [doi: [10.1037/hea0000305](https://doi.org/10.1037/hea0000305)] [Medline: [26651463](https://pubmed.ncbi.nlm.nih.gov/26651463/)]

33. Coughlin LN, Nahum-Shani I, Philyaw-Kotov ML, Bonar EE, Rabbi M, Klasnja P, et al. Developing an adaptive mobile intervention to address risky substance use among adolescents and emerging adults: usability study. *JMIR Mhealth Uhealth* 2021 Jan 15;9(1):e24424 [FREE Full text] [doi: [10.2196/24424](https://doi.org/10.2196/24424)] [Medline: [33448931](https://pubmed.ncbi.nlm.nih.gov/33448931/)]
34. Beyer F, Lynch E, Kaner E. Brief interventions in primary care: an evidence overview of practitioner and digital intervention programmes. *Curr Addict Rep* 2018 May 3;5(2):265-273 [FREE Full text] [doi: [10.1007/s40429-018-0198-7](https://doi.org/10.1007/s40429-018-0198-7)] [Medline: [29963364](https://pubmed.ncbi.nlm.nih.gov/29963364/)]

Abbreviations

CIAS: Computerized Intervention Authoring System

COREQ: Consolidated Criteria for Reporting Qualitative Studies

MI: motivational interviewing

Edited by A Mavragani; submitted 15.10.21; peer-reviewed by B Chaudhry, MDG Pimentel; comments to author 03.11.21; revised version received 09.12.21; accepted 30.12.21; published 30.03.22.

Please cite as:

Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR

The Perceived Benefits of Digital Interventions for Behavioral Health: Qualitative Interview Study

J Med Internet Res 2022;24(3):e34300

URL: <https://www.jmir.org/2022/3/e34300>

doi: [10.2196/34300](https://doi.org/10.2196/34300)

PMID: [35353045](https://pubmed.ncbi.nlm.nih.gov/35353045/)

©Gabriela Marcu, Steven J Ondersma, Allison N Spiller, Brianna M Broderick, Reema Kadri, Lorraine R Buis. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 30.03.2022. 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.

Original Paper

Barriers and Considerations in the Design and Implementation of Digital Behavioral Interventions: Qualitative Analysis

Gabriela Marcu¹, PhD; Steven J Ondersma², PhD; Allison N Spiller¹, MSI; Brianna M Broderick¹, MSI; Reema Kadri³, MLIS; Lorraine R Buis^{1,3}, PhD

¹School of Information, University of Michigan, Ann Arbor, MI, United States

²Department of Obstetrics, Gynecology, & Reproductive Biology and the Division of Public Health, Michigan State University, East Lansing, MI, United States

³Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:

Lorraine R Buis, PhD

Department of Family Medicine

University of Michigan

1018 Fuller Street

Ann Arbor, MI, 48104-1213

United States

Phone: 1 734 998 7120

Fax: 1 734 998 7335

Email: buisl@umich.edu

Abstract

Background: Digital behavioral interventions have become increasingly popular for their ability to support patient diagnosis and treatment, chronic disease self-management, behavior change, and adherence to recommended care. However, digital intervention development is impeded by challenges such as limited technical skills, limited access to developers, and cost. The purpose of this study is to elicit in-depth qualitative feedback from intervention developers who have interest in digital behavioral interventions but lack programming skills regarding the barriers they experience and key considerations in the design and implementation of digital interventions.

Objective: This study aims to understand barriers in the design and implementation of digital behavioral interventions, as well as to identify key considerations for researchers who are developing these interventions.

Methods: We conducted semistructured qualitative interviews with 18 researchers who had experience either designing (but not coding) digital behavioral interventions or running research studies with them. Participants were a convenience sample of users of the Computerized Intervention Authoring System platform, an existing no-code development platform for building digital intervention content, and were recruited through either direct email solicitation or snowball sampling. All interviews were conducted and recorded over videoconference between February and April 2020. Recordings from interviews were transcribed and thematically analyzed by multiple coders.

Results: Interviews were completed with 18 participants and lasted between 24 and 65 (mean 46.9, SD 11.3) minutes. Interviewees were predominantly female (17/18, 94%) and represented different job roles, ranging from researcher to project/study staff. Three key barriers in the development of digital behavior interventions were identified during interviews: lack of cross-disciplinary understanding; variability in recipients' technology access, infrastructure, and literacy; and the idea that evidence-based in-person interactions do not translate directly to digital interactions. Interviewees identified several key considerations that interventionists learned to prioritize, which have the potential to overcome these barriers and lead to successful interventions.

Conclusions: Barriers in the development of digital behavioral interventions are often created by a lack of cross-disciplinary understanding, which can lead to difficulties conceptualizing interventions, unrealistic expectations in terms of cost, and confusion about the development process. Moreover, concerns about research study participant characteristics and access to technology, as well as the translation of in-person interventions to digital, are apparent. Appropriate training in how to work with software development teams may help future digital behavior intervention creators overcome these barriers and may lead to new, exciting innovations in this space.

(*J Med Internet Res* 2022;24(3):e34301) doi:[10.2196/34301](https://doi.org/10.2196/34301)

KEYWORDS

computers; mobile apps; screening; brief interventions; diagnosis; computer-assisted/methods; surveys and questionnaires; motivational interviewing; therapy; implementation; qualitative; mobile phone

Introduction

Background

Digital behavioral interventions have become increasingly popular for their ability to support patient diagnosis and treatment, chronic disease self-management, behavior change and adherence to recommended care, and primary prevention [1]. They have demonstrated broad promise in terms of efficacy [2-8], but evidence for specific use cases tends to be mixed [9-11], with engagement and retention being particular challenges [9,12-14], as well as low-quality evidence [10,15,16].

There is tremendous room for innovation in this area. However, developing digital interventions is often out of reach for many research teams because of the challenges of developing relationships with software developers, the costs of custom software development, and the need for technical expertise, all of which serve as barriers to entry in this field. Critically, researchers need pilot data to access grant funding large enough to develop these types of interventions but may not be able to collect pilot data without at least a working prototype. Moreover, the constant need to update and adapt custom tools for other purposes means that research teams are constantly in the process of reinventing the wheel with each new intervention. Finally, most digital interventions are built using a particular technology stack by a particular group of developers, such that sharing or building on existing interventions is a significant challenge, even with the few apps that make their source code openly available.

No-code platforms, also known as authoring tools, may offer a solution to these challenges. Part of what has been called the no-code revolution, these platforms are designed to allow citizen developers [17] to build apps using simple graphical user interfaces. By providing templates and other structured step-by-step processes for building custom interventions, these platforms enable the creation of technology-delivered interventions without the need for computer programming or technical knowledge. For example, an increasing number of services now enable anyone to create their own website, including relatively complex functions such as e-commerce, responsive web design, and analytics. No-code platforms could similarly make the development of digital interventions faster, more accessible, and easier to edit and share by providing a framework within which researchers can build or tailor an intervention. A successful platform could be to the creation of digital behavioral interventions what Microsoft PowerPoint is to the creation of slides for a presentation: a tool facilitating an explosion of content generation.

In contrast, as with Microsoft PowerPoint, democratization of software development does not solve all challenges in the creation, testing, and implementation of digital behavioral interventions [17]. In this study, we sought to explore these challenges, grounded in the experiences of researchers who have played a range of roles with digital interventions. The

purpose of this 2-part qualitative investigation is to better understand the motivations, needs, constraints, and experiences of researchers who have looked to digital interventions in their work addressing behavioral health.

Objectives

In this paper (part 2), we seek to document common barriers reported by intervention developers, as well as key considerations, in the design and implementation of digital interventions. In a companion paper (part 1), we seek to document reasons why researchers who study behavioral health focus their efforts on digital interventions, as well as their perspectives on the perceived benefits that digital approaches afford researchers and their intervention recipients [18].

Methods

Overview and Ethics Approval

This study was conducted as part of a redesign of the Computerized Intervention Authoring System (CIAS) platform, a web-based digital behavioral intervention authoring tool for researchers. This study included previous and current CIAS users as well as CIAS-naïve users, all of whom had experience developing behavioral health interventions. We conducted semistructured interviews with these intervention creators to better understand their needs, perceived barriers to designing and implementing digital interventions, and design considerations for behavioral intervention development. COREQ (Consolidated Criteria for Reporting Qualitative Studies) guidelines were followed [19]. All methods used in this study were approved by the University of Michigan Human Subjects Review Board (HUM00171197) and Wayne State University (IRB-19-10-1340).

CIAS Platform

CIAS is a Health Insurance Portability and Accountability Act-compliant no-code web app designed to allow users to easily build, edit, and share web-based interventions without coding or other technical expertise of any kind. CIAS gives interventionists the ability to develop tailored and personalized text-based interventions that can be narrated by an animated and emotive character capable of multiple different actions and voices (supporting >40 different languages, with male and female versions of most and a range of accents or dialects for some of the more commonly spoken languages). CIAS supports intervention building features such as multiple question types, natural language reflections, branching and tailoring, and integration of video content. Further details on CIAS can be found in the part 1 companion paper [18]. With support from the National Institutes of Health (EB028990), an all-new version of CIAS (3.0) is currently being developed as an open-source and noncommercial research resource. This research was conducted as part of the CIAS 3.0 redesign activities.

Participant Recruitment

Participants were recruited via email solicitation from a convenience sample of CIAS users and researchers who have expressed interest in using CIAS, via snowball sampling, or via an email to a listserv of University of Michigan Department of Family Medicine staff members familiar with conducting digital interventions. To be eligible to participate, users were required to be aged at least 18 years.

Study Procedures and Data Collection

All interviews were conducted in single sessions via teleconference between February and April 2020 by 2 trained research staff (BMB and ANS). Participants did not have prior experience with the interviewers, nor did they have access to interview guides prior but knew that they would be participating in redesign efforts. All participants provided verbal consent, and all interviews were recorded for later transcription. Interviewers also took field notes, which were available for later analysis. Interviews were conducted until saturation was reached. Transcripts were not returned to participants for review, nor were participants asked to provide feedback regarding the findings. All participants were offered US \$20 for their participation, which was delivered in the form of a check via US mail.

Analysis

To analyze the data, the 2 study team members (BMB and ANS) debriefed after each interview. During this debrief, individual notes were compared, which were gradually synthesized across interviews. Interview transcripts underwent 2 rounds of inductive thematic analysis, which was conducted by 2 coders (BMB and ANS); the first focused on identifying patterns, which later became themes, and then a second validated themes and confirmed connections between them.

Results

Overview

We invited 24 current and former CIAS users to participate, and 17 agreed (17/24, 71% response rate). In addition, we recruited 1 additional CIAS-naïve user who was familiar with digital health interventions from the University of Michigan Department of Family Medicine. For our 18 interviews, the average duration was between 24 and 65 (mean 46.9, SD 11.3) minutes. Participants were predominantly female (17/18, 94%) and represented different job roles ranging from researcher to project or study staff. [Table 1](#) presents the aggregate interviewee characteristics, and [Table 2](#) presents the brief descriptions of each interviewee.

Table 1. Aggregate interviewee characteristics (N=18).

Characteristics	Values, n (%)
Gender	
Male	1 (6)
Female	17 (94)
Job title (all that apply)	
Researcher	13 (72)
Psychologist	4 (22)
Other (project manager or coordinator and research assistant or associate)	5 (28)
Employer	
Academic institution	15 (83)
Foundation	1 (6)
Contract research organization	1 (6)
Other (nonprofit research organization)	1 (6)
Race	
African American or Black	4 (22)
White	13 (72)
Prefer not to answer	1 (6)
Ethnicity	
Hispanic or Latino	2 (11)
Non-Hispanic or Non-Latino	16 (89)
Education	
Bachelor's	6 (33)
Master's	4 (22)
Beyond a master's degree	8 (44)
Self-reported proficiency with CIAS^a (n=17)	
Novice	3 (18)
Proficient	4 (24)
Advanced	7 (41)
Expert	3 (18)

^aCIAS: Computerized Intervention Authoring System.

Table 2. Individual interviewee descriptions.

ID	Self-reported title	Employer type	Highest level of education obtained	Gender
P1	Researcher; psychologist	Academic institution	Beyond a master's degree	Female
P2	Researcher	Foundation	Master's degree	Female
P3	Project manager	Academic institution	Bachelor's degree	Female
P4	Psychologist	Academic institution	Beyond a master's degree	Female
P5	Project manager	Academic institution	Master's degree	Female
P6	Researcher	Academic institution	Beyond a master's degree	Male
P7	Researcher	Academic institution	Beyond a master's degree	Female
P8	Researcher; psychologist	Academic institution	Beyond a master's degree	Female
P9	Researcher; psychologist	Academic institution	Beyond a master's degree	Female
P10	Researcher	Academic institution	Bachelor's degree	Female
P11	Researcher	Academic institution	Bachelor's degree	Female
P12	Research assistant	Academic institution	Bachelor's degree	Female
P13	Researcher	Contract research organization	Master's degree	Female
P14	Researcher; research project coordinator	Academic institution	Master's degree	Female
P15	Researcher	Nonprofit research organization	Beyond a master's degree	Female
P16	Research associate	Academic institution	Bachelor's degree	Female
P17	Researcher	Academic institution	Bachelor's degree	Female
P18	Researcher	Academic institution	Beyond a master's degree	Female

Barriers in Designing and Implementing Digital Behavioral Interventions

Interviews revealed three primary barriers: lack of cross-disciplinary understanding; variability in recipients' technology access, infrastructure, and literacy; and evidence-based in-person interactions do not translate directly to digital interactions.

Barrier 1: Lack of Cross-disciplinary Understanding

The first barrier commonly described by interviewees was the technical knowledge and resources required to use digital behavioral interventions. For some, this was the primary barrier: "I think the biggest barrier for many researchers is the actual building of the software itself, or how the intervention will be delivered" (P7). Building even a simple digital intervention requires the appropriate technical expertise and project management skills. Consequently, a significant amount of effort must often come from software developers who are outside the clinical research team, which increases the cost. The resources required were described by our interviewees as prohibitive, as well as difficult to anticipate:

The cost of actually developing an app is, I think, prohibitive for a lot of researchers....I think it depends also on your resources and the amount of time you have, you know, we have a small pilot grant from NIH, so we don't have a lot of resources and we don't have a lot of time and I think, I mean, I certainly didn't know going in...it just seems like everything, every piece of this, takes longer than we anticipated and the programming is a lot more intricate. [P15]

The bounds of a clinical team's technical knowledge may limit their ability to not only implement their ideas through technology but also to even envision what possibilities are available to them:

I think it's just a lack of knowledge around technology. Like, what options are available. I think that the go-to thought process when people think digital interventions is telehealth...but there's a lot more to that. So, I think just knowing that there are other options is a barrier. A lot of people are limited in that. But then once they realize that there are other options...it's accessibility to these things. Like, what is out there? How much does it cost? How do I do it? You know, a lot of the people designing these interventions aren't...they don't have a background in technology. [P16]

Conversely, those with technical knowledge who develop tools are often limited in their understanding of how interventions should be delivered. Interviewees discussed the challenges of using digital interventions that had been created without input from clinical experts:

The people who are developing these interventions are not people who deliver them. So that's a problem, I think, because sometimes it's really hard to envision what should it really look like in the development stage when you're not the kind of person who would be sitting across the table from someone, right? So like, you can't figure out what the gap is because how would you know? Because you've never been sitting across the table from someone. [P1]

As a result, interviewees found that existing software programs had significant limitations in how they could be used for developing and delivering interventions. For example, they did not allow the interviewee to edit and control important aspects of an intervention: “I think the lack of editability of the program, I think that’s the biggest thing that’s frustrating for me” (P12). Interviewees therefore indicated the need to truly combine knowledge of behavioral interventions with knowledge of technology that can deliver them. However, those who had worked closely with software developers noted the challenges inherent in this cross-disciplinary work:

That can be a barrier because oftentimes there’s disciplinary differences that challenge that experience, you know, speaking different language, computer programmers not quite “getting” behavioral intervention and the needs of the behavioral interventionist. [P7]

This interviewee also went on to explain that the level of investment in technology involves trade-offs that are commonly discussed and debated among researchers:

Particularly when we’re at conferences and seeing the work of other people, is that process of working with software engineers and developing software from the ground up is not a short process...I’ve seen animated programs that are really fun and interactive and interesting, but they also took 2 years to build...So, we’ve kind of talked about that trade-off before. Then you think about the fact that most people stop using an app or web-based type of a program, typically in a short period of time. You’ve got that huge investment in that fancy intervention software and then people are using it for a short period of time and never looking at it again, and that’s a huge investment of resources for what could be a limited return and an unknown efficacy. [P7]

The level of resources required to develop a digital intervention is a large investment and one that comes with some risk. Therefore, setting realistic expectations for timeline, cost, capabilities, and the return on investment is necessary before turning to digital interventions. One interviewee explained how they share their past experience to help other researchers adjust their expectations:

So, if they say “oh, we want a long course [for the intervention], we have multiple interventions, we have \$5,000” something like that. I will just tell them that \$5,000 is not enough, we need a lot more than that, however, no matter what you have, you can usually get something, but it may not be what you want. So, it’s \$5,000—it’s going to be crude messaging, it’s going to be texts, it’s going to be tailored audio tracks, or mailed—simple feedback. You could do that, but you’re not going to do anything fancy with video, or anything like that, for a couple thousand dollars. I guess that’s the last thing, that’s helping people to budget what kind of money they have, what kind of expectations are. [P6]

Planning and budgeting an intervention with realistic expectations can also help interventionists to think through how presentation of the content will affect intervention efficacy. Some interviewees shared their concern that the look of the digital intervention is linked to its credibility and perceived value: “Better optics might make participants take it more seriously” (P12). Interviewees therefore felt that investing in the aesthetics of the intervention could even impact engagement and efficacy.

Finally, we note that interviewees mentioned a set of technical barriers after the software development had been largely completed. As interventionists finalized, pilot-tested, and prepared an intervention for recipients, adjustments and edits still needed to be done within the code. For example, changing the terminology used on a button may seem straightforward; however, because the label on the button is hard coded into its functionality, the interventionists will not be able to change it themselves. The process of communicating every single change, large or small, and waiting for all of them to be completed can significantly extend project timelines. One interviewee described this extensive process and how experiencing it changed the way they would approach writing grants by accounting for each step in the process:

The communication [with the software development company] and the ability to sort of get something corrected just adds time to it relative to things that we could otherwise have done in-house...We would submit what we needed. They came back with things they could do easily, things that would take more time, things that would cost more money. And so that was probably several months of just sort of figuring out what was going to be feasible given the limitations of grant funding. And then probably another couple of months of back and forth finalizing items and then the student pilot tested with a couple of participants and worked out a few more glitches before we launched...But that whole process, you know again not having really done that with another company...just made me appreciate kind of, if I had to write another grant with that company how I might do it. [P9]

Digital behavioral interventions require significant knowledge from multiple disciplines, which naturally raises costs and increases the complexity of project communication and timelines. Through experience with digital interventions, interventionists get better at envisioning what they can do with technology, having appropriate expectations for a software development project, and managing costs.

Barrier 2: Variability in Recipients’ Technology Access, Infrastructure, and Literacy

Technical barriers from the recipients’ side include access to technologies such as mobile devices or Wi-Fi, literacy with these technologies, and the availability and cost of local infrastructure such as internet connection and data plans. These factors are often underestimated when digital interventions are being designed and planned and also lead to greater resources that need to be contributed by interventionists, such as supplying

Wi-Fi, data plans, or additional technical support. In addition, as operating systems (eg, Android [Google] and iOS [Apple Inc]) are updated over time, digital interventions may require recipients to install these updates on devices to ensure their continued operation.

Interviewees referenced the digital divide, the significant disparity between those who have access to technology and have fluency with technology and those who do not. Most commonly, interviewees mentioned mobile devices such as tablets (eg, iPads) and smartphones, as these are the primary technologies currently used for delivering behavioral interventions:

Not everybody has a computer. Not everybody has an iPad at home. Most—a lot of people do have smartphones, but even then, like it doesn't mean that they're going to be able to access the information or they're willing to access the information. [P2]

In order for technology to effectively facilitate a behavioral intervention, recipients must have not only access to the right device but also the appropriate technology literacy to operate and troubleshoot the device. Interviewees experienced the digital divide between themselves and the people they were trying to reach, acknowledging that the decision to use a digital intervention was a much easier one based on their own considerations compared with those for recipients:

For us, it was an easy idea for us to, you know, be like “yea, we'll do it on an iPad! No big deal.” But then, in talking to the mothers that we're potentially going to be using this behavioral intervention with, we quickly realized that a lot of these participants have probably never held an iPad in their hands before and that kind of totally changed the way we need to develop our intervention to where it's, not only are you creating your intervention, but you have to create the instructions and infrastructure in place so that they can even experience the intervention in the way that you think it's going to go. [P13]

In ensuring the right experience for intervention recipients, this interviewee alluded to the instructions needing to match their technology literacy, a consideration very similar to health literacy. With regard to infrastructure, this can include the type of internet connection accessible to a recipient in their neighborhood, as well as the data plan they have on their phones:

With our population...I know everyone has a phone but a lot of [the] women use prepaid phones with minutes. And they have...a limit on how much data they can use. [P16]

Often, the varied technological constraints can be underestimated and may not be considered when digital interventions are chosen, designed, and initially implemented:

The extent of it came up organically, like just how much we kind of took for granted when it came to technological literacy with these [participants]. I think we didn't realize that we would have to, in some cases, provide our own Wi-Fi, and that we can't just ask these [participants] for an email address to get

their feedback report, so things like that came up organically, the whole idea that we needed to know where these people were at with technology. [P13]

Interventionists also did not always consider compatibility issues based on which versions of hardware or software recipients were used. For example, an older tablet may cause more issues or not be able to run the intervention. As operating systems evolve with updates regularly released, users must download and install these updates on their devices; otherwise, the intervention may no longer run. When an intervention stops running as expected, the cause is not always obvious to the user, and they may need help troubleshooting. Although these challenges are par for the course with any software project, interventionists may not be prepared for them:

All of a sudden, we're dealing with hot spots, and intermittent access, and old computers, and updates. The versions, what we do now, Android will update a version and all of a sudden it will break, it won't work. So, our people in the field were like, “oh, it's not working, people are sitting around, the clients are getting frustrated,” and your tech people will say “we don't know what's happening, it works fine on our—,” you know. Computers are way more difficult, so those are some of the reasons people don't do it is because they don't know how. [P6]

Interviewees had learned how much extra work was required from their research team when technical challenges arose for the recipients:

For this new clinical trial, when they're going to be doing it by themselves, like it's already hard enough to get people to enroll in these studies and then for it to not work properly. And then I'm, they're going to reach out to me, my project manager, like, “well, this is not working.” And then it makes more work for them, okay, go take a screenshot of this point where it didn't work and then email it to me. It's like, oh it's a whole production. [P12]

Interviewees were therefore concerned about the effects on both the recipients and research staff. Moreover, technical difficulties with a digital intervention could stand in the way of effective behavior change, jeopardizing impact on the recipients, as well as the course of research:

I'm a little bit concerned about potential technical difficulties that they might experience that might lead to the intervention not being completed, which would then limit our ability to test the efficacy of the intervention. [P7]

Designing an intervention that meaningfully engages a recipient in the process of behavior change is challenging enough on its own, but a digital intervention adds multiple levels of technical considerations.

Barrier 3: Evidence-Based In-Person Interactions Do Not Translate Directly to Digital Interactions

The final barrier we found concerns the shift interventionists had to make from designing an in-person intervention to designing a digital intervention. Interventionists look to

technology for many perceived advantages [18]; however, interviewees pointed to the difficulty of adapting person-delivered interventions to digital forms. For example, despite the capability to program a digital narrator in CIAS so that it can adjust content in reaction to input from the user, interventionists still missed the richness of other cues they were used to relying on during in-person interactions:

Trying to translate the patient-provider interaction and the types of feedback and information that goes into that—how can you program that? [The digital narrator] is pretty much 100% scripted, he doesn't really do a lot of responding to cues. [P7]

The CIAS platform allows interventions to be delivered through a digital narrator whose voice reads out the intervention text and whose physical form appears in the interface as a human or animal. In this physical form, the digital narrator can be animated for a range of actions, such as talking, smiling, reading, waving, and pointing to items within the interface. The narrator can also ask recipients for their preferences, reactions, and thoughts at any point and can provide verbal reflections of that input (eg, “It sounds like this has been pretty hard for you”) and branch to different areas of content based on that input. However, interventions built using the CIAS platform cannot read facial expressions or body language and cannot interpret unscripted input from recipients, leading to challenges in forecasting or collecting a range of likely reactions or preferences so they can be included in the intervention:

I think part of the challenge that we've had is we've often trying to take what we would do in a face-to-face and think about how that would translate to a computerized delivery, which there's obviously going to be elements of that that are lost with the computer delivery. [The narrator] can squawk and flap his wings and make cute expressions, but he can't really interpret the non-verbal patient sitting across from him so to speak...When you're in the face-to-face setting, you obviously can use non-verbal and other types of communication to inform your thinking: looking at patients' records, looking at other sources of information, downloading meters, and other things that can be conversation start points that [the narrator], he can't necessarily do. [P7]

Although digital interventions built with CIAS allow for personalization, tailored feedback, and empathic reflections, the need for a priori entry of possible preferences and reactions is a departure from the flexibility and improvisation that is possible with person-delivered interventions:

I guess the biggest problem that I would say is that, you know, you're just in a forced choice kind of framework...So like, for example...if I were to ask you “hey, what do you think would...what would help you supervise your kid every day?” So, let's say you...rate yourself as motivated but low on self-efficacy...so, if I'm sitting across the table from you, I can just say to you, “well, what are some things you've tried?” or “what do you think would work?” and we can have an open-ended conversation. Whereas in the program,

as the developer, I have to presuppose what your likely barriers are, and put them in there, so you can click some boxes. [P1]

Interventionists, therefore, experienced greater difficulty when they approached the design of a digital intervention in the same way as a traditional human-to-human intervention. Some described the need instead for computational thinking or thinking of an intervention based on the characteristics and advantages of computers, such as logic and chunking:

But the difficult thing is to bridge that gap between what can be done in person versus what can be done online. What helps me is to think through sort of logical rules or how would you chunk things. And there are an infinite number of possibilities of what could happen in an in-person interaction. [P6]

However, as interventionists drew on behavioral theory and approaches, such as motivational interviewing (MI) [20], their expertise in applying them was primarily via human interaction, making it difficult to envision a different type of intervention. For example, interviewees pointed out that the ability to incorporate recipient responses into the course of the intervention was a key component of MI:

The most challenging thing is taking the human part out and trying to put that into mobile platforms...so, for example, like I said with MI, branching out the reflections and making it so that you could take someone's motivation and put it on a scale...that was challenging versus asking someone “hey, can I know more about this?”...I think that's a challenge...because as a human being, you could just adapt to [their response]. [P5]

The loss of nuanced reactions they were used to observing in person also meant fewer opportunities for gathering rich qualitative data:

I think, with the regular behavioral interventions that I work on it's a lot more communication with participants and like face-to-face interaction which is kind of nice because you're getting in like real-time reactions and kind of some good qualitative data. But I guess with [the digital intervention] they're doing on their own you're not kind of seeing their reaction to the actual intervention. [P3]

Human contact is, of course, not easily replicated with technology, which some interviewees saw as a significant loss:

So, the biggest thing that I would consider is, a lot of participants that I've seen in the past, the number one thing that they get out of being involved in research is that they get to talk to somebody, they get that communication, they get that personal connection. So, like if there's a chance that they're going to—in their eyes—lose that, just the way that you approach how you're going to transition it. Like, we're not going to be losing contact, we're going to still be in contact, and how you're going to be in contact. Because a lot of people that we deal with, they have severe depression and anxiety and so they participate

in these things obviously to try and you know, reduce those symptoms. But also, because they don't have anybody at home so they like seeing somebody in person. [P11]

This interviewee highlighted the potential benefits to intervention recipients, especially when working with sensitive topics and vulnerable populations, as a key consideration when choosing a digital intervention. Combining digital intervention with strategically positioned human touchpoints may also be an effective way to maintain such benefits for recipients.

However, integrating digital interventions smoothly into the experience and constraints of a recipient or an interventionist may benefit from new approaches. When designing a digital intervention, we found that interviewees often work to translate traditional interventions directly, maintaining a very similar model or approach that they had previously used in person. For example, one intervention relied on an avatar, whose role was viewed similarly to that of a human counselor helping the recipient to navigate the steps of the intervention:

We have a little [avatar of a] woman that walks you through the whole process. She kind of acts as the counselor for the participant and talks them through everything, and we build scripts for her. Kind of, build like, questions and answers and get to know the participant through CIAS. [P12]

This interviewee discussed the process of writing a script for the avatar, much like a script might be generated to support consistent human delivery of an intervention. However, achieving a flow similar to that of a natural conversation was a challenge with this approach:

I think at times it tends to, and this might be on the side of the script building and our side of putting CIAS together in the back end, but I've noticed at times that it tends to lag or, kind of lack the flow, it needs to kind of seem like a conversation. [P12]

Digital interventions need not be designed according to the traditions and best practices that have been developed with face-to-face interventions. Instead, digital interventions may be improved by envisioning new paradigms.

Key Considerations for Designing and Implementing Digital Behavioral Interventions

Overview

As interviewees shared the barriers and lessons learned from past experiences with various interventions and recipients, we identified three high-level considerations for designing and implementing digital behavioral interventions: understanding the population and context, integrating the intervention within ecologies of care, and technical staffing and preparation. A strong focus on these considerations can help address the barriers we have described previously.

Understanding the Population and Context

As discussed earlier, the process of adapting behavioral approaches to digital interactions with recipients is one of the greatest barriers to effective implementation of digital

interventions. Therefore, the first step in understanding the technological context of recipients is critical not only for ensuring appropriate access and literacy to enable intervention delivery but also for envisioning how each step of the intervention will be experienced by recipients, so that its delivery is evidence based and recipient centered. Interviewees gained this understanding through qualitative methods with the population and frontline interventionists who interact directly with the population (eg, via home visits):

In our first focus groups, we were talking to both home visitors and moms, [asking]: How comfortable are you using technology? If someone handed you an iPad and said "go through this program on your own," how comfortable would you be doing that? What are some things that would get in the way of you being able to do that? And then asking the home visitors, "when you go into a home, is there Wi-Fi? Because this is a web-based program." Things like that. And then we had them sit down and go through the program and analyze both their experience with it, just as the platform in general, and then also nitty-gritty of the wording of the content and what we're trying to accomplish. [P13]

This interviewee highlighted some of the best practices in designing digital interventions: starting the process by engaging with the population through a method such as focus groups to gather information pertinent to their experience and assessing each aspect of the intervention through their actual perspective, from the experience of interacting with the technology to clarity of the content presented. Another best practice mentioned was pilot testing, also known as usability testing, which examines how the recipient interacts with the intervention to identify how that experience might be improved:

We plan to do a small, pilot trial to test it out in home visiting, and because it's a pilot we consider [it] part of the design process where we're really looking at feasibility and acceptability, which parts work and which don't, and then we'll continue to tweak it from there. [P15]

As this interviewee points out, there must be an iterative process of testing and revising. As such, obtaining input on aspects of the design from end users should occur from the early stages of designing a digital intervention. Deeper engagement, known as participatory or community based, can involve the population in not only evaluating the intervention but also contributing ideas to its design:

I think everyone needs to do more of a community-based approach, if they can, to where they're involving the target population in the creation of [the digital intervention]. [P13]

Engaging the population in evaluation, or even design, can help address a range of concerns about how the intervention will be received. For example, interviewees mentioned that when it comes to the overall aesthetic of the intervention, "the way it looks is how seriously someone is going to take your research regardless of the content" (P14). Interviewees also made comparisons to the aesthetics of common mobile apps, such as

social media apps. Referring to a prior version of CIAS, one interviewee said:

It could just look a little bit more, you know, like the social media apps that everybody uses all the time. And like if we saw a social media app that looked like that, you'd be like "ew, what is this app, like it looks weird." I love CIAS, not to say that I hate it, I appreciate it, but I'd like a graphic redesign of CIAS to make it a little bit more appealing to the eye. [P12]

Such natural comparisons are important to consider because how an app looks and feels to a user and how the intervention content is presented can be linked to perceived value or trustworthiness of the intervention as well as the research study.

Finally, digital interventions expand the possible contexts in which recipients can be reached, changing the way they must be designed. Deciding on the geographical constraints for reaching a certain population has important implications for the intervention, and interviewees described this decision as a balance between convenience for the recipients and convenience for the research team:

Do you want to restrict it to certain geographical areas...what's your population?...Where are these people located, the time zones? Because we have people like in California. We have to consider you know, because we have to do follow up calls every six weeks after they've completed their treatment and sometimes we have to consider that they're three hours behind us, so what time might be good for them, we have to accommodate, because they are doing us a favor by participating in our study so if that means that we have to call them at 8:00 pm our time that's fine....So those are the kind of barriers, it just really depends on how widespread you want your remote study to be. [P11]

Interventionists took into consideration a range of factors about the population they wanted to reach with a digital intervention, working to understand their background, technological context, how they would interact with the digital intervention, what their expectations may be of interactive technologies generally, and how geographically distributed they may be.

Integrating the Intervention Within Ecologies of Care

Interventionists also considered how an intervention might be integrated within ecologies of care—existing services and supports, both in virtual and real-life contexts. Our interviews indicated that researchers have a range of ideas for integration but have been constrained in achieving all of the features they envision. For example, they thought recipients should be able to schedule a health appointment and even see a clinician at a distance using the same or another platform:

Having a feature that if a client wants to have a digital health appointment for something, they could set that up on their own in the [digital intervention]. They're scheduling it and it's happening on the platform. [P16]

Another potential feature mentioned was supporting patients to monitor their progress toward goals, which can include reminders, strategically timed prompts to action, or visual feedback on what it will take to meet their goals. One interviewee described the use of this approach for long-term engagement:

Have your [digital interventions] just send them text messages: "hey you set a goal for whatever, just make sure you do that today." That's the cheapest and easiest way [to improve patient compliance]. If you can tie your [digital interventions] in [with] some long-term prodding of your patients in the next week or two or three or whatever, in a way that clinicians think they're going to reap that. If the patient has been prompted to do something, the clinician thinks: "gee, the next time they come in, maybe they'll have gotten their material together and put it in the binder and developed a diet, talked to a nutritionist, or visited a grocery store, or found out more about where they can get nicotine replacement therapy." So all those I think are selling points. [P6]

Similarly, care transitions, such as leaving residential or inpatient care, are a great challenge in which technologies could provide support. Interviewees described roles for digital interventions across multiple touchpoints in one's journey with treatment for substance use:

I think I could see [digital interventions] particularly serving a role in transitions from things like residential to outpatient treatment. That's that window I think we're all still really struggling with is, how do you get [recipients] to maintain abstinence when they leave residential care? Where I think trying to build things in with brief interventions would be a great place. So, I think within the addiction treatment realm of specialized services I would see [digital interventions] as something that could help with engagement at the front end of things—to provide some early counseling in a way that facilitates retention in treatment and even enrollment in some cases. And then I think in the longer term [digital interventions] could be used for interventions that help to reduce relapse risk and facilitate uses of care. [P9]

Our interview findings suggest that even if the technology alone cannot provide meaningful interaction, the way it is integrated into other touchpoints of the research study itself can facilitate human contact that may be able to make a greater impact:

I don't think [the digital aspect of the intervention] detracts from our interaction because we still have to get to know each other over the year that they're in the study, so talking with them over the phone, and asking them really personal questions like adverse events form and things like that about their lives. [P12]

Conversely, the way the technology is used can help minimize human interaction for the purpose of protecting privacy when engaging recipients on sensitive topics:

We also use headphones. So that's something that we do to make sure that it's anonymous, because the whole idea is that people don't want to talk about these issues with their home visitors. So we're giving them an iPad so that it's just them and the iPad. [P2]

When choosing a digital intervention, it is important to remember that recipients are not likely to interact only with the technology. Integrating the digital intervention effectively within the structure of the research study, as well as the broader landscape of services and supports, will enable a good experience for recipients and may even improve impact or outcomes.

Technical Staffing and Preparation

Finalizing, testing, and preparing a digital intervention to ensure that it is ready for use is a complex process with various considerations. Our interviews revealed the extent to which research teams need to have staff available to focus on each aspect of this preparation. For example, once scripts for the avatar were written, multiple members of the research team listened to how they would be delivered by the automated voice to ensure a good experience for the recipients:

[The PI] did a lot of the scripting, like, actually what the character would say and then she and I both would put it in there and listen to it and see how it flows...I think that we really worked hard to make the language understandable and modern. We didn't use a lot of very complicated words. It is an automated voice, so we really played around with like how words sound. [P16]

The flow of audio and visual features also needed to be reviewed on the same device and in the same way that recipients would experience them. CIAS enables interventionists to design their intervention on a laptop or desktop computer while previewing what the intervention will look like (eg, on the smaller screen of a mobile device). However, such previews will not convey the full user experience, for example, leaving out touch screen interactions, which will affect how intuitively and efficiently a recipient is able to navigate through the intervention. Therefore, researchers would ensure to review the intervention on the actual device their recipients would be using:

And then I think the last step is just making sure that it looks nice on the iPad. So, just going through it as many times and listening to it. ... If I'm previewing [the intervention] on my laptop, it's not going to be the same on an iPad. [P2]

Interviewees also noted that a digital intervention may not look or perform the same across different platforms, such as tablets versus smartphones or Android versus Apple. These potential differences had implications for testing across platforms, as well as members of the research team having enough familiarity with operating different platforms so that they can troubleshoot during a study:

I have really tailored our intervention to a tablet screen. I haven't looked at it on a phone or you know...on a computer in a long time. So, I like the cross platform capability to be able to like take

something that's working on one and make sure it's going to look the same or be usable on another device—something I'm actively trying to look at right now. [P16]

I have never used any digital intervention on an Android phone, so I'm not that comfortable [troubleshooting potential problems for recipients]. I would assume that it works the same way, but I don't know. Or just things like, [the recipient's] device functionality not being the same as ours [Apple device]. [P12]

In addition to testing the intervention on an appropriate device, interventionists needed to test it on an internet connection within the community that was similar to how the recipients would be accessing the intervention. This was especially important in communities or contexts in which internet connections may be unreliable or restricted (eg, behind log-ins, paywalls, or firewalls):

I'm going to ask my [research assistants] to go out into the community, whether that's a library or to a Starbucks, or wherever and see if they can run through the intervention without getting kicked out. [P7]

One interviewee worked as a project manager and spent a significant amount of time testing all branches of the intervention logic. As she suggests, CIAS or other types of platforms could potentially provide support specifically for this type of testing:

Where I have the most experience was doing the beta testing before we could actually run it...you had to test every single branch because there is no way within CIAS to like click a button to see the flow. [P5]

Testing did not end once a study had begun. Research teams would have to ensure an adequate internet connection for individual recipients, who may be distributed across different contexts. This activity was so critical that it had been formalized as a part of study protocols:

In our protocol, we have outlined different ways to kind of go about that barrier [of having Wi-Fi]. Our participants can be all over [the state] or all over the US. They could be in places where there is just not Wi-Fi. So, our first thing is suggesting they go to a public place that has free Wi-Fi and we can help them figure that out. Is that a bookstore, or is that a little coffee shop or something, and then if they can't, then we'll direct them to our IT person and possibly work with them on getting them a mobile hotspot. [P11]

If Wi-Fi could not be guaranteed, some study teams carried data cards in the field, that is, prepaid cards that can be inserted into a mobile device for internet connectivity without Wi-Fi. The data cards were used for either a recipient's device if it did not have a data plan that could support the intervention or the research team's device when it was more cost-effective than purchasing a data plan. The data cards added to the research assistants' responsibilities for materials and troubleshooting:

So we have the data cards and we have the iPad for the site study I have actually been doing the digital intervention with. The upcoming one will be [on] their device, but still I would bring the [data] card and they would still use our iPad to do questionnaires. [P12]

The approach of providing technology for recipients was a common way to overcome the digital divide and ensure a more consistent experience for recipients. This strategy still requires sufficient staffing on the research team to provide technical support, but troubleshooting becomes significantly more straightforward. For example, one research team distributed Kindle e-readers to recipients:

We knew that technology was going to be a downfall so we have a tech personnel on call to answer any questions that need to be....They are receiving a Kindle, so if they don't have a computer, they don't have a phone, it's okay, they have the Kindle, and they get to keep the Kindle. We modify it so that it's very easy, so on the home screen it's like this is your app. [P11]

Compared with asking recipients to use their own devices, supplying devices eliminates the need to provide instructions or support for downloading because the research team can ensure that study devices reach recipients with the app already installed on them. Some research teams also allow recipients to keep study devices, either as an added incentive or to avoid the challenging logistics of collecting the devices and reusing them at the end of the study.

Many digital interventions require recipients to have email addresses so that they can set up profiles or accounts. However, research teams cannot assume that all recipients would already have their own email address, and interviewees needed to plan for research assistants to take the additional step of helping to create one when needed:

If [the recipients] don't have an email address, my research assistant works with them and she helps them create an email address just for them. [P11]

Although support for creating an email address is easy to provide, this illustrates the need to be nimble and prepared to troubleshoot and assist, which can require a significant amount of time from research staff.

Finally, interviewees described taking the time to set recipient expectations for how they will experience the intervention. This activity was described not as a training but as an orientation or a demo, indicating that the technology itself should be intuitive to use, but recipients still ought to know what to expect as with any research activity:

We're gonna talk with the participant during what we're calling "intervention orientation," where we kind of explain kind of how the intervention is going to be delivered and how it's going to work. [P7]

We've demoed it on a computer or on an iPad through a projector so that way people can see what it looks like. I've also gone in and created a short video...of what people can expect to see throughout the

intervention, because I think it's important for people when they're receiving it to know what they can expect. So like, "this is what questions will look like."... "this is what a link will look like, and this is what will happen if you click on a link." [P2]

After a research team has put considerable thought into designing their intervention and invested in the complex process of building the intervention according to their design, the key phase of ensuring that their intervention will be implemented effectively with recipients remains.

Discussion

Principal Findings

Our findings from these researchers with experience in digital intervention development are 2-fold. First, we isolated three common barriers: lack of cross-disciplinary understanding; variability in recipients' technology access, infrastructure, and literacy; and evidence-based in-person interactions do not translate directly to digital interactions.

Second, we identified three key considerations that interventionists had learned to prioritize across different types of interventions and projects: understanding the population and context, integrating the intervention within ecologies of care, and technical staffing and preparation. These considerations are important from the outset of project planning and budgeting, and focusing on them can help address the barriers. We revisit each barrier to contextualize these findings and build on what interviewees had offered as their key considerations to using digital interventions, by adding our own recommendations.

Barriers in Designing and Implementing Digital Behavioral Interventions

Lack of Cross-disciplinary Understanding

Although digital behavioral interventions are often developed by multidisciplinary teams, the nascent ideas for these programs tend to start with either clinicians or researchers who do not have experience with software development or with developers with technical backgrounds but no clinical experience. As our interviewees were interventionists and not technical developers, themes that arose from this study centered on the lack of cross-disciplinary understanding that results when nontechnical interventionists seek to develop digital behavioral interventions.

Conceptualizing Digital Behavioral Interventions

Interviewees expressed concern about the lack of technical expertise on the part of intervention creators and the effects that this may have on interventions. The ideas for digital behavioral interventions conceptualized by clinical researchers are often built on the shoulders of other interventions. For example, a researcher may see an interesting intervention that applies to 1 disease state and brainstorm ways in which it could be adapted or replicated for use in another disease state. Although this approach is common, when it is used by nontechnical clinical researchers, the new idea may unnecessarily pigeonhole new interventions by ignoring different intervention delivery modalities (eg, websites, native mobile apps, web apps, and smartphone apps), design features (eg, manual self-monitoring

and pairing with peripheral devices for objective monitoring), mechanisms for communicating content (eg, text, audio, video, and animation), or design decisions that the creator may not have considered or even be aware of.

Misaligned Expectations of Development Costs

Emerging technologies often remain out of reach for many interventionists because of their cost, which depending on factors such as complexity, novelty, and timeline, may vary greatly. Therefore, it is crucial for interventionists to align their expectations of what they want and need to develop, with the resources available to support intervention development. Misalignment with wants or needs and available resources often stem from a lack of experience with digital behavioral intervention development or the costs necessary to run and maintain an intervention, such as expenses related to hosting, security, bug fixes, and other maintenance costs. Most behavioral interventions used within research need to be implemented reliably at a scale within certain budgetary constraints. Interviewees voiced many concerns related to development costs, indicating that funds available for a given intervention development project were often far less than what was actually needed. Indeed, budget caps on federal and private research caps are often incompatible with actual market costs for high-quality software development and maintenance, which may stifle innovation. In particular, development costs are often greatest for new interventions, which are least likely to be competitive for large grant awards. Interventionists interested in advancing the state of the art in digital interventions often need to partner with technology researchers on more sophisticated and innovative interventions, often at a smaller scale, greater cost, and limited immediate ability to demonstrate impact on behavior.

Working With Software Developers

Interventionists often lack the knowledge and skills required to build digital interventions themselves and must partner with software development companies to create apps; however, doing so requires preparation that most researchers lack, at least initially. For example, even the process of identifying the right software development company can be a challenge. Some clinical researchers expressed confusion around their options when it came to hiring developers, saying that they did not understand what they should look for or how to know who to hire. In addition, before an intervention can be developed, the clinical researcher must have a clear vision and proposal for the intervention, including full details of each question, how it should be presented digitally, and an outline of all functionality. From hardware to software to the user interface, there are intricate details to a digital intervention that are time-consuming to design and will affect the quality of the intervention. Many of these details require careful decisions that researchers have not previously considered. Interviewees described lack of knowledge about what was even possible. Finally, developing and testing the intervention takes time and iteration. The processes, steps, and timelines imposed by working with a software development company can be quite lengthy and costly, growing more as researchers need even small changes made to a developed intervention. Working with outside development companies can create a disciplinary gap because those building

the intervention are not typically knowledgeable about interventions. This can lead to more back-and-forth between clinical researchers and developers to be on the same page.

Understanding the Product Design Cycle and Development Timeline

Clinical researchers who do not have first-hand experience with the product design cycle used within software development may not know how to plan their projects around it. If they do not have realistic expectations, they can then find themselves with timeline issues such as delays and budget problems associated with not locking in functional requirements and specifications in a timely manner, difficulties in making system changes because of an advanced stage of coding, or failure to include necessary components or design processes into contractual language. These can all result in digital health tools that miss the mark both for the creators who envisioned the interventions and for end users. Moreover, many digital behavioral intervention projects are behind schedule from the moment they are funded because of unanticipated contracting delays. Ensuring adequate communication with the clinical researcher, who in this case is the client for which the software is being developed, is not typically the role of a software developer, programmer, or engineer. More commonly, the responsibility for understanding the needs of the client and end user of the software lies with consultants, product managers, or user experience researchers. These types of roles specialize in overcoming cross-disciplinary communication challenges and translating between technical language and the language of, in this case, the behavioral interventionist. Unfortunately, our interviews suggest that the prohibitive costs of software development may lead researchers to work directly with those who will write the code, asking them to take on a broad range of roles that are critical to a successful software project. Working with an experienced third-party vendor, or investing in these other roles and specializations in-house, can make cross-disciplinary collaboration more effective for creating an impactful intervention, and even thoughtfully target the most difficult problems, such as engagement with technology over time.

Meeting Recipient Expectations for Digital Technology

Interviewees shared their concerns about ensuring that their intervention aligned with expectations that their recipients have for digital technologies they use in their everyday lives. Given the ubiquity of high-quality free apps and mobile websites and the technology industry's investment in the user experience of their products, consumers have high expectations for the appearance and functionality of interventions. The most prominent apps in their lives (eg, for social media, e-commerce, email, and search) are created by companies that invest considerably in the user experience, with entire departments dedicated to evaluating and continually improving the experience for their users. Apps that do not provide an experience that users have come to expect can generate disappointment, lower engagement, or abandonment.

The concerns of our interviewees echo the problem of engagement, which is well established in the literature [9,12-14,21,22] and which may be exacerbated when digital

health tools do not provide a high-quality user experience. Interventionists should therefore draw from industry best practices to meet their recipients' expectations. Asking end users of a digital intervention to interact with it, and assessing their experience, is an essential best practice [23] for ensuring the technology is user-centered and intuitive to use. For example, mockups of how the digital intervention will look can be shown in a study by Melles et al [23] or as a static image on a tablet. Users can then be asked to think aloud [24] as they review the information presented, react to the information, and describe what actions they believe they can take next. Such qualitative methods are the most effective [23] to ensure that the digital intervention is designed in a way that matches the mental model of the population. Achieving an adequate user experience can be costly, not necessarily because this requires complexity (in fact, current technology trends favor minimalism), but by investing in a user experience designer or interaction designer who will know how to meet user needs and expectations, including up-to-date standards and conventions to make it feel modern and credible. Moreover, engaging with the population as various aspects of the intervention are being designed can reveal how different design choices will affect the way they experience the intervention, providing opportunities to correct the course as needed while the intervention is still taking shape. These activities add further complexity to the timelines and budgets of software development projects we previously discussed. However, these activities are commonly adapted to fit the constraints of each project and can even save time and money when used early and strategically.

Variability in Recipients' Technology Access, Infrastructure, and Literacy

Interviewees in this study expressed many concerns about the technology access, infrastructure, and literacy of target end users, and these issues need to be considered by digital behavioral health interventionists, particularly those that work with underserved or marginalized populations. Although access to computers and smartphones may be high, access is not ubiquitous in all populations. Moreover, technology infrastructure is not equitable across the country, particularly in rural and underresourced urban settings where access to affordable, dependable, high-quality, high-speed broadband may not be readily available for all [25,26]. Furthermore, intervention-generated inequalities may arise where already advantaged populations gain greater advantage through digital interventions than do those who are less advantaged [27].

Finally, technology literacy is required to ensure that a device is connected to the internet, any peripheral devices such as a keyboard or wearable technology are paired, effective security measures such as passwords are used, and software updates are downloaded and installed in a timely manner to keep the device or intervention running smoothly. This requires a fair amount of technology literacy not only on the part of end users but also on the study staff assisting research participants to ensure adequate assistance with troubleshooting when problems arise. Regardless of the amount of advanced planning and testing, unanticipated issues with the technology will arise with any intervention, and these will affect recipients and study staff. Consequently, interventionists need to have a mechanism for

receiving reports of technical problems or difficulties. Troubleshooting problems, assisting recipients, and communicating with developers require additional effort and expertise from interventionists. Given that the interventionists perform less of the intervention delivery, their role therefore shifts more to technical training and support for recipients and ongoing communication with developers.

Evidence-Based In-Person Interactions Do Not Translate Directly to Digital Interactions

Overview

Our interviews revealed the extent to which interventionists approached the creation of digital interventions in similar ways as they had in the past with traditional methods and how they could run into challenges as a result. Although this is related to the conceptualization of the design of an intervention, which we addressed in an earlier section, we see it as a broader mindset and therefore discuss it separately. We highlight that moving analog processes and behavioral interventions to a digital world is not a 1:1 replication, nor should it be. That which is efficient and effective offline may not translate to the digital world in the same way, and the benefit of digitization means that there are potential efficiencies that can be captured with redesigned processes that are not possible in an analog world. This applies both to the front end of an intervention that is visible to research participants and also to the back-end databases and structures that run the interventions. For example, one can certainly take print materials and data collection forms and put them in a digital format as is, but this may not be the most efficient and effective content delivery solution. Moreover, given that many behavioral interventions are delivered in a face-to-face format with trained interventionists using techniques such as MI, the switch to digital platforms, which are often automated, cannot be a perfect replication. Interviewees from this study raised several concerns about the translation of in-person interventions and interactions to a digital world, including issues related to intervention design and intervention fidelity.

Intervention Design

Despite access and literacy variability among intervention recipients, one of the potential benefits valued by researchers is the ability to reach more people through the use of digital interventions at scale [18]; however, this extended reach of digital interventions comes with added complexity for planning and implementing interventions. In addition to time zone considerations, there may be other regional or cultural factors that affect how the intervention will be experienced by recipients across geographic areas. An early understanding of the potential differences between populations in different areas will enable effective design and planning. In addition, the need for tech-savviness on the part of intervention recipients introduces a level of complexity that may be off-putting for individuals less comfortable with technology. This increases the need to introduce experts in human-computer interaction and interaction design, who specialize in developing conceptual models through which users can intuitively interact with technology (eg, the desktop on a computer interface is derived from the metaphor of a physical desktop with files and folders placed upon it). In contrast to supporting intervention planning by outlining a range

of functional requirements to consider [28], focusing on user-centered conceptual models can help interventionists pivot from how they think of traditional interventions to how they might conceptualize digital interventions differently. Although some interventions try to emulate human counselors in order to provide an approach that mirrors real life, models other than that of the human counselor should be explored to effectively guide recipients through an intervention and to conceptualize how digital interventions can complement in-person services. Previous work has shown that although intervention recipients can develop a bond with an app and display open communication using this medium, the nature of the interaction is different than that with health professionals, as apps are not as able to simulate truly human qualities such as friendliness and collaboration [29]. Indeed, simulating realistic exchanges between humans and computers within the context of health issues has the potential to feel to human users as disingenuous, contrived, or overly stereotypical, which may be a deterrent to use [30]. There is evidence that suggests SBIRT (Screening, Brief Intervention, and Referral to Treatment) interventions can be at least as good as face-to-face, clinician-delivered interventions [31,32].

Interviewees often talked about the loss of ability to gauge and react to intervention recipients' responses, behaviors, and body language. The ability of automated interventions to handle such nuanced verbal and nonverbal communication is still limited; however, we note that artificial intelligence and other emerging technologies do have the potential to improve personalization and adaptivity of interventions through speech pattern recognition, natural language processing, and machine learning. These advanced technological methods can help to process the large number of possible intervention pathways and make decisions based on certain rules that are preprogrammed or learned over time, addressing the challenge articulated by some interviewees. These innovations are developing and improving rapidly and mark an important area of study at this time. Furthermore, as interventionists look to digital interventions to supplement human interaction and support, there is a critical need for continued investigation of the recipient experience. Evidence is needed to determine the extent to which technology

can convey empathy and emotional support and, if so, through which features or functions specifically. Finally, it should also be noted that although technology clearly cannot do everything that a human interventionist can do, neither can a human interventionist do everything that technology can do, and those advantages should be leveraged freely, rather than only seeking to replicate human interaction. The literature in this area includes examples of technology-delivered interventions performing equally well as human-delivered interventions [31], with similar overall acceptability [33] and greater cost-effectiveness [34].

Limitations

This study was limited, first, by our focus on the CIAS platform, given that this work was conducted within the context of CIAS redesign activities. Although the limitation on CIAS users may have potentially introduced selection and response bias, the issues raised by interviewees in this study were largely focused on considerations for developing and implementing digital behavioral interventions in general, and not on CIAS itself. Our participants described experiences across a range of behavioral interventions. We also contextualized our recommendations within the broader literature and best practices for software projects. Finally, our sample size was small; however, saturation was achieved, leaving us confident that adding additional participants to the sample would yield little new information to contribute.

Conclusions

Researchers from within the health and social sciences who wish to develop and implement digital behavioral interventions describe a consistent set of challenges. Issues resulting from lack of cross-disciplinary understanding; variability in recipients' technology access, infrastructure, and literacy; and translating face-to-face interventions into digital interactions are common across multidisciplinary teams in this space and speak to the need for better training and planning among research teams who wish to work with digital health. Finding the right expertise to include on multidisciplinary teams can help teams overcome these challenges and build off similar work that has been done, as opposed to reinventing the wheel with every new project.

Acknowledgments

This research was funded by the National Institute of Biomedical Imaging and Bioengineering at the National Institutes of Health under grant number EB028990 to SJO. The authors would like to thank Lilly Pritula and Rania Ajilat for administrative support, including manuscript editing, and the interviewees who made this study possible.

Conflicts of Interest

None declared.

References

1. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *J Med Internet Res* 2017;19(6):e232 [FREE Full text] [doi: [10.2196/jmir.7126](https://doi.org/10.2196/jmir.7126)] [Medline: [28663162](https://pubmed.ncbi.nlm.nih.gov/28663162/)]
2. Du S, Liu W, Cai S, Hu Y, Dong J. The efficacy of e-health in the self-management of chronic low back pain: a meta analysis. *Int J Nurs Stud* 2020;106:103507. [doi: [10.1016/j.ijnurstu.2019.103507](https://doi.org/10.1016/j.ijnurstu.2019.103507)] [Medline: [32320936](https://pubmed.ncbi.nlm.nih.gov/32320936/)]

3. Kaner EF, Beyer FR, Garnett C, Crane D, Brown J, Muirhead C, et al. Personalised digital interventions for reducing hazardous and harmful alcohol consumption in community-dwelling populations. *Cochrane Database Syst Rev* 2017;9(9):CD011479 [FREE Full text] [doi: [10.1002/14651858.CD011479.pub2](https://doi.org/10.1002/14651858.CD011479.pub2)] [Medline: [28944453](https://pubmed.ncbi.nlm.nih.gov/28944453/)]
4. Kaner EF, Beyer FR, Muirhead C, Campbell F, Pienaar ED, Bertholet N, et al. Effectiveness of brief alcohol interventions in primary care populations. *Cochrane Database Syst Rev* 2018;2(2):CD004148 [FREE Full text] [doi: [10.1002/14651858.CD004148.pub4](https://doi.org/10.1002/14651858.CD004148.pub4)] [Medline: [29476653](https://pubmed.ncbi.nlm.nih.gov/29476653/)]
5. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. *PLoS One* 2017;12(3):e0173160 [FREE Full text] [doi: [10.1371/journal.pone.0173160](https://doi.org/10.1371/journal.pone.0173160)] [Medline: [28249025](https://pubmed.ncbi.nlm.nih.gov/28249025/)]
6. Sin J, Galeazzi G, McGregor E, Collom J, Taylor A, Barrett B, et al. Digital interventions for screening and treating common mental disorders or symptoms of common mental illness in adults: systematic review and meta-analysis. *J Med Internet Res* 2020;22(9):e20581 [FREE Full text] [doi: [10.2196/20581](https://doi.org/10.2196/20581)] [Medline: [32876577](https://pubmed.ncbi.nlm.nih.gov/32876577/)]
7. Xu H, Long H. The effect of smartphone app-based interventions for patients with hypertension: systematic review and meta-analysis. *JMIR Mhealth Uhealth* 2020;8(10):e21759 [FREE Full text] [doi: [10.2196/21759](https://doi.org/10.2196/21759)] [Medline: [33074161](https://pubmed.ncbi.nlm.nih.gov/33074161/)]
8. Zhai YK, Zhu WJ, Cai YL, Sun DX, Zhao J. Clinical- and cost-effectiveness of telemedicine in type 2 diabetes mellitus: a systematic review and meta-analysis. *Medicine (Baltimore)* 2014;93(28):e312 [FREE Full text] [doi: [10.1097/MD.0000000000000312](https://doi.org/10.1097/MD.0000000000000312)] [Medline: [25526482](https://pubmed.ncbi.nlm.nih.gov/25526482/)]
9. Beleigoli AM, Andrade AQ, Caçado AG, Paulo MN, Diniz MF, 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;21(1):e298 [FREE Full text] [doi: [10.2196/jmir.9609](https://doi.org/10.2196/jmir.9609)] [Medline: [30622090](https://pubmed.ncbi.nlm.nih.gov/30622090/)]
10. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual research review: digital health interventions for children and young people with mental health problems - a systematic and meta-review. *J Child Psychol Psychiatry* 2017;58(4):474-503. [doi: [10.1111/jcpp.12663](https://doi.org/10.1111/jcpp.12663)] [Medline: [27943285](https://pubmed.ncbi.nlm.nih.gov/27943285/)]
11. Shaw G, Whelan ME, Armitage LC, Roberts N, Farmer AJ. Are COPD self-management mobile applications effective? A systematic review and meta-analysis. *NPJ Prim Care Respir Med* 2020;30(1):11. [doi: [10.1038/s41533-020-0167-1](https://doi.org/10.1038/s41533-020-0167-1)] [Medline: [32238810](https://pubmed.ncbi.nlm.nih.gov/32238810/)]
12. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. *J Med Internet Res* 2019;21(9):e14567 [FREE Full text] [doi: [10.2196/14567](https://doi.org/10.2196/14567)] [Medline: [31573916](https://pubmed.ncbi.nlm.nih.gov/31573916/)]
13. Bennett GG, Glasgow RE. The delivery of public health interventions via the Internet: actualizing their potential. *Annu Rev Public Health* 2009;30:273-292. [doi: [10.1146/annurev.publhealth.031308.100235](https://doi.org/10.1146/annurev.publhealth.031308.100235)] [Medline: [19296777](https://pubmed.ncbi.nlm.nih.gov/19296777/)]
14. Stockings EA. Commentary on Boumparis et al. (2017): it works, but only if they use it-addressing treatment adherence and attrition in internet interventions for illicit substance use. *Addiction* 2017;112(9):1533-1534. [doi: [10.1111/add.13862](https://doi.org/10.1111/add.13862)] [Medline: [28778118](https://pubmed.ncbi.nlm.nih.gov/28778118/)]
15. Janjua S, Banchoff E, Threapleton CJ, Prigmore S, Fletcher J, Disler RT. Digital interventions for the management of chronic obstructive pulmonary disease. *Cochrane Database Syst Rev* 2021;4(4):CD013246. [doi: [10.1002/14651858.CD013246.pub2](https://doi.org/10.1002/14651858.CD013246.pub2)] [Medline: [33871065](https://pubmed.ncbi.nlm.nih.gov/33871065/)]
16. Thabrew H, Stasiak K, Hetrick SE, Wong S, Huss JH, Merry SN. E-Health interventions for anxiety and depression in children and adolescents with long-term physical conditions. *Cochrane Database Syst Rev* 2018;8(8):CD012489 [FREE Full text] [doi: [10.1002/14651858.CD012489.pub2](https://doi.org/10.1002/14651858.CD012489.pub2)] [Medline: [30110718](https://pubmed.ncbi.nlm.nih.gov/30110718/)]
17. Invernizzi P, Tossell B. Decoding the no-code/low-code startup universe and its players. Medium. URL: <https://pinver.medium.com/decoding-the-no-code-low-code-startup-universe-and-its-players-4b5e0221d58b> [accessed 2021-10-10]
18. Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR. The perceived benefits of digital interventions for behavioral health: qualitative interview study. *J Med Internet Res* 2022;24(3):34300 [FREE Full text] [doi: [10.2196/34300](https://doi.org/10.2196/34300)]
19. 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;19(6):349-357. [doi: [10.1093/intqhc/mzm042](https://doi.org/10.1093/intqhc/mzm042)] [Medline: [17872937](https://pubmed.ncbi.nlm.nih.gov/17872937/)]
20. Miller WR, Rollnick S. Motivational interviewing: helping people change. 3rd edition. New York: Guilford Press; 2012.
21. Eysenbach G. The law of attrition. *J Med Internet Res* 2005;7(1):e11. [doi: [10.2196/jmir.7.1.e11](https://doi.org/10.2196/jmir.7.1.e11)] [Medline: [15829473](https://pubmed.ncbi.nlm.nih.gov/15829473/)]
22. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. *J Med Internet Res* 2010;12(4):e69 [FREE Full text] [doi: [10.2196/jmir.1640](https://doi.org/10.2196/jmir.1640)] [Medline: [21156470](https://pubmed.ncbi.nlm.nih.gov/21156470/)]
23. Melles M, Albayrak A, Goossens R. Innovating health care: key characteristics of human-centered design. *Int J Qual Health Care* 2021;33(Supplement_1):37-44 [FREE Full text] [doi: [10.1093/intqhc/mzaa127](https://doi.org/10.1093/intqhc/mzaa127)] [Medline: [33068104](https://pubmed.ncbi.nlm.nih.gov/33068104/)]
24. Fan M, Shi S, Truong KN. Practices and challenges of using think-aloud protocols in industry: an international survey. *J Usability Stud* 2020;15(2):85-102. [doi: [10.1145/1753326.1753685](https://doi.org/10.1145/1753326.1753685)]
25. About a quarter of rural Americans say access to high-speed internet is a major problem. Pew Research Center. 2018. URL: <https://www.pewresearch.org/fact-tank/2018/09/10/about-a-quarter-of-rural-americans-say-access-to-high-speed-internet-is-a-major-problem/> [accessed 2021-10-10]

26. Internet/broadband fact sheet. Pew Research Center. 2021. URL: <https://www.pewresearch.org/internet/fact-sheet/internet-broadband/?menuItem=c41259a2-d3a8-480d-9d1b-2fb16bcf0584> [accessed 2021-10-10]
27. Veinot TC, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. *J Am Med Inform Assoc* 2018;25(8):1080-1088 [FREE Full text] [doi: [10.1093/jamia/ocy052](https://doi.org/10.1093/jamia/ocy052)] [Medline: [29788380](https://pubmed.ncbi.nlm.nih.gov/29788380/)]
28. Cunha BC, Rodrigues KR, Zaine I, da Silva EA, Viel CC, Pimentel MD. Experience sampling and programmed intervention method and system for planning, authoring, and deploying mobile health interventions: design and case reports. *J Med Internet Res* 2021;23(7):e24278 [FREE Full text] [doi: [10.2196/24278](https://doi.org/10.2196/24278)] [Medline: [34255652](https://pubmed.ncbi.nlm.nih.gov/34255652/)]
29. Berry K, Salter A, Morris R, James S, Bucci S. Assessing therapeutic alliance in the context of mHealth interventions for mental health problems: development of the mobile Agnew Relationship Measure (mARM) questionnaire. *J Med Internet Res* 2018;20(4):e90 [FREE Full text] [doi: [10.2196/jmir.8252](https://doi.org/10.2196/jmir.8252)] [Medline: [29674307](https://pubmed.ncbi.nlm.nih.gov/29674307/)]
30. Gilbey D, Morgan H, Lin A, Perry Y. Effectiveness, acceptability, and feasibility of digital health interventions for LGBTIQ+ young people: systematic review. *J Med Internet Res* 2020;22(12):e20158. [doi: [10.2196/20158](https://doi.org/10.2196/20158)] [Medline: [33270039](https://pubmed.ncbi.nlm.nih.gov/33270039/)]
31. Martino S, Ondersma SJ, Forray A, Olmstead TA, Gilstad-Hayden K, Howell HB, et al. A randomized controlled trial of screening and brief interventions for substance misuse in reproductive health. *Am J Obstet Gynecol* 2018;218(3):322.e1-322.12 [FREE Full text] [doi: [10.1016/j.ajog.2017.12.005](https://doi.org/10.1016/j.ajog.2017.12.005)] [Medline: [29247636](https://pubmed.ncbi.nlm.nih.gov/29247636/)]
32. Yonkers KA, Dailey JI, Gilstad-Hayden K, Ondersma SJ, Forray A, Olmstead TA, et al. Abstinence outcomes among women in reproductive health centers administered clinician or electronic brief interventions. *J Subst Abuse Treat* 2020;113:107995 [FREE Full text] [doi: [10.1016/j.jsat.2020.02.012](https://doi.org/10.1016/j.jsat.2020.02.012)] [Medline: [32359666](https://pubmed.ncbi.nlm.nih.gov/32359666/)]
33. Loree AM, Yonkers KA, Ondersma SJ, Gilstad-Hayden K, Martino S. Comparing satisfaction, alliance and intervention components in electronically delivered and in-person brief interventions for substance use among childbearing-aged women. *J Subst Abuse Treat* 2019;99:1-7 [FREE Full text] [doi: [10.1016/j.jsat.2019.01.007](https://doi.org/10.1016/j.jsat.2019.01.007)] [Medline: [30797381](https://pubmed.ncbi.nlm.nih.gov/30797381/)]
34. Olmstead TA, Yonkers KA, Ondersma SJ, Forray A, Gilstad-Hayden K, Martino S. Cost-effectiveness of electronic- and clinician-delivered screening, brief intervention and referral to treatment for women in reproductive health centers. *Addiction* 2019;114(9):1659-1669 [FREE Full text] [doi: [10.1111/add.14668](https://doi.org/10.1111/add.14668)] [Medline: [31111591](https://pubmed.ncbi.nlm.nih.gov/31111591/)]

Abbreviations

CIAS: Computerized Intervention Authoring System

COREQ: Consolidated Criteria for Reporting Qualitative Studies

MI: motivational interviewing

SBIRT: Screening, Brief Intervention, and Referral to Treatment

Edited by A Mavragani; submitted 15.10.21; peer-reviewed by B Chaudhry, MDG Pimentel; comments to author 05.11.21; revised version received 09.12.21; accepted 30.12.21; published 30.03.22.

Please cite as:

Marcu G, Ondersma SJ, Spiller AN, Broderick BM, Kadri R, Buis LR

Barriers and Considerations in the Design and Implementation of Digital Behavioral Interventions: Qualitative Analysis

J Med Internet Res 2022;24(3):e34301

URL: <https://www.jmir.org/2022/3/e34301>

doi: [10.2196/34301](https://doi.org/10.2196/34301)

PMID: [35353043](https://pubmed.ncbi.nlm.nih.gov/35353043/)

©Gabriela Marcu, Steven J Ondersma, Allison N Spiller, Brianna M Broderick, Reema Kadri, Lorraine R Buis. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 30.03.2022. 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.

Original Paper

Social-Cyber Maneuvers During the COVID-19 Vaccine Initial Rollout: Content Analysis of Tweets

Janice T Blane¹, BSc, MSc; Daniele Bellutta¹, BSc, MSc; Kathleen M Carley¹, BSc, PhD

School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, United States

Corresponding Author:

Janice T Blane, BSc, MSc
School of Computer Science
Carnegie Mellon University
5000 Forbes Ave
Pittsburgh, PA, 15213
United States
Phone: 1 412 268 6016
Fax: 1 412 268 1744
Email: jblane@andrew.cmu.edu

Abstract

Background: During the time surrounding the approval and initial distribution of Pfizer-BioNTech's COVID-19 vaccine, large numbers of social media users took to using their platforms to voice opinions on the vaccine. They formed pro- and anti-vaccination groups toward the purpose of influencing behaviors to vaccinate or not to vaccinate. The methods of persuasion and manipulation for convincing audiences online can be characterized under a framework for social-cyber maneuvers known as the BEND maneuvers. Previous studies have been conducted on the spread of COVID-19 vaccine disinformation. However, these previous studies lacked comparative analyses over time on both community stances and the competing techniques of manipulating both the narrative and network structure to persuade target audiences.

Objective: This study aimed to understand community response to vaccination by dividing Twitter data from the initial Pfizer-BioNTech COVID-19 vaccine rollout into pro-vaccine and anti-vaccine stances, identifying key actors and groups, and evaluating how the different communities use social-cyber maneuvers, or BEND maneuvers, to influence their target audiences and the network as a whole.

Methods: COVID-19 Twitter vaccine data were collected using the Twitter application programming interface (API) for 1-week periods before, during, and 6 weeks after the initial Pfizer-BioNTech rollout (December 2020 to January 2021). Bot identifications and linguistic cues were derived for users and tweets, respectively, to use as metrics for evaluating social-cyber maneuvers. Organization Risk Analyzer (ORA)-PRO software was then used to separate the vaccine data into pro-vaccine and anti-vaccine communities and to facilitate identification of key actors, groups, and BEND maneuvers for a comparative analysis between each community and the entire network.

Results: Both the pro-vaccine and anti-vaccine communities used combinations of the 16 BEND maneuvers to persuade their target audiences of their particular stances. Our analysis showed how each side attempted to build its own community while simultaneously narrowing and neglecting the opposing community. Pro-vaccine users primarily used positive maneuvers such as excite and explain messages to encourage vaccination and backed leaders within their group. In contrast, anti-vaccine users relied on negative maneuvers to dismay and distort messages with narratives on side effects and death and attempted to neutralize the effectiveness of the leaders within the pro-vaccine community. Furthermore, nuking through platform policies showed to be effective in reducing the size of the anti-vaccine online community and the quantity of anti-vaccine messages.

Conclusions: Social media continues to be a domain for manipulating beliefs and ideas. These conversations can ultimately lead to real-world actions such as to vaccinate or not to vaccinate against COVID-19. Moreover, social media policies should be further explored as an effective means for curbing disinformation and misinformation online.

(*J Med Internet Res* 2022;24(3):e34040) doi:[10.2196/34040](https://doi.org/10.2196/34040)

KEYWORDS

social cybersecurity; social-cyber maneuvers; social network analysis; disinformation; BEND maneuvers; COVID-19; coronavirus; social media; vaccine; anti-vaccine; pro-vaccine; ORA-PRO; cybersecurity; security; Twitter; community; communication; health information; manipulation; belief

Introduction

Background & Motivation

COVID-19 claimed the lives of 2.6 million people in the first year of its discovery [1]. In a concerted effort to reduce the cases and deaths resulting from the COVID-19 pandemic, governments and major health organizations pushed for the development and rapid distribution of COVID-19 vaccines. This process, however, has been met with online expressions of resistance vaccination [2]. In view of this concerning spread of anti-vaccine sentiment online, our work has focused on identifying the specific tactics used by both the pro- and anti-vaccine communities to spread their messages over Twitter.

Though vaccinating everyone against COVID-19 may seem to be an obvious way to prevent deaths, many people and groups oppose vaccination for several different reasons. The first compulsory vaccination was established in England by the Vaccination Act of 1853. The act faced opposition to the idea that the government should impose health legislation [3]. In current times, communities speak out against the government and assert that they have the right to decide what goes inside their bodies. Some anti-vaccine proponents fear the side effects of vaccines and refuse entirely to vaccinate themselves or their children because of rumors of autism or other medical disorders [4,5].

The Pfizer-BioNTech vaccine was the first vaccine for preventing COVID-19 to be authorized in the United States by the Federal Drug Administration (FDA). Both the FDA and European Medicines Agency (EMA) authorized the vaccine for emergency use. In 2020, the first Pfizer vaccine doses were distributed in the United Kingdom on December 8 and in the United States on December 14. Because of the rush to create the vaccine, many feel the vaccines were inadequately tested and refuse the vaccine without seeing the results of long-term studies. Additionally, some accept conspiracy theories or rumors on the vaccine. For example, one such conspiracy theory is that Bill Gates and the government created the vaccines to microchip the population for some malicious intent [6]. These are among the reasons for “vaccine hesitancy” across the world [7].

Social media has become a medium for COVID-19 vaccine discussion. Twitter is a popular platform on which government leaders, public health officials, and news organizations spread pertinent information. However, many users spread misinformation or act maliciously by conducting influence campaigns to manipulate peoples’ beliefs and ideas. Bonnevie et al [2] found that vaccine opposition on Twitter increased by 80% after COVID-19 began spreading in the United States. Misinformation is not limited to anti-vaccine users as some pro-vaccine users also share unreliable information [8]. To counter the spread of misinformation on its platform during the initial administration of the vaccine, Twitter expanded its policy by removing false and misleading tweets about COVID-19

vaccines, adding labels to potentially misleading COVID-19 vaccine information, and creating a “five-strike system” for suspending misleading accounts [9,10].

These malicious actions online are a major aspect within the field of social cybersecurity. Social cybersecurity lies at the intersection between cyberspace and human interaction. It studies how humans can be influenced by tactful messaging and connecting the right people to the right content. Key players in an online social network can conduct influence maneuvers to change users’ beliefs and affect their behavior [11]. In our study, we sought to identify the important actors in pro- and anti-vaccine Twitter communities as well as the social-cyber maneuvers they used to influence their audiences’ stances regarding the COVID-19 vaccine.

This work focused on the time period around the approval and initial administration of the Pfizer vaccine. Our objective was to determine whether there are differences between the types of social-cyber maneuvers pro-vaccine and anti-vaccine communities use toward their target audiences. We describe a methodology for determining pro-vaccine or anti-vaccine stances within tweets and identifying key players within the social network. We further used bot detection and linguistic cues to analyze the content and significance of tweets, and we evaluated how the 2 opposing vaccine communities applied social-cyber maneuvers to persuade their target audiences. Our results show how pro-vaccine messaging focused on exciting readers and explaining the vaccine issue. In contrast, anti-vaccine groups preferred to make dismaying statements and used messaging that distorted vaccine information. We also found that Twitter’s tightening of its policies on vaccine misinformation had a remarkable effect on decreasing the size of anti-vaccine communities and the prevalence of their messaging.

Related Work

Vaccine Stance Detection

The problem of identifying pro- and anti-vaccine communities has garnered the attention of several researchers who have sought to apply stance detection techniques from computer science to this task. Supervised machine learning methods developed for this problem have ranged from the use of transformer neural networks based on Google’s Bidirectional Encoder Representations from Transformers (BERT) model [12] to the use of convolutional neural networks trained on n-grams and topics detected via Latent Dirichlet Allocation [13]. More traditional community detection algorithms have also been used to find groups with overt stances on vaccines [14]. The semisupervised stance propagation technique used for our work, which has the advantage of not requiring extensive manual labelling of pro- and anti-vaccine messages, was also used to identify linguistic differences between pro- and anti-vaccine groups [15].

Pro-Vaccine and Anti-Vaccine Communities

Various studies of anti-vaccine and pro-vaccine communities have sought to identify the methods used for spreading vaccine-related messages. Different communities can have contrasting messaging characteristics depending on the nature of and support for their stances on vaccines.

In 2019, a study examining influential themes and actors within the anti-vaccine community concluded that top tweeters relied on highly networked communities led by accounts that select messages expected to have high receptivity within those communities [16]. This was different from standard messages from public officials, which tended to repeat the same information to the same communities, limiting the extent of a message's reach. In an analysis of Facebook vaccine group clusters, Johnson et al [17] observed that anti-vaccination clusters entangled more often with undecided clusters, while pro-vaccination clusters tended to be more peripheral. Furthermore, Schmidt et al [14] examined how echo chambers reinforce the opinions of groups and how involvement within these groups could be an effective way of countering anti-vaccine beliefs.

Past research has therefore found that pro-vaccine messages tend to be supported by public health officials and governments seeking to reduce the spread of infectious diseases, whereas anti-vaccine communities are more niche and maintain a smaller following. However, although pro-vaccine messages tend to stay within pro-vaccine communities, anti-vaccine messages permeate beyond the boundaries of anti-vaccine communities.

Although these past works have analyzed the themes and targeting of vaccine messaging, they have not considered the specific types of strategies carried out in vaccine-related information operations. Thelwall et al [18] tracked some of the anti-vaccine narratives spreading on Twitter, and Boucher et al [19] identified the key themes in Twitter conversations about vaccine hesitancy. However, previous research has not examined the intentions behind specific choices on the language, content, and targeting of pro- and anti-vaccine messaging. Our work breaks down the tactical value of specific types of vaccine messages and analyzes how those tactics have changed over time.

Social Cybersecurity: Influence Campaigns and Bots

A key development in the fight against online influence campaigns has been the growth of the field of social cybersecurity, a computational social science that aims to protect the security of democratic societies by studying the ways in which actors exercise manipulation on social media platforms [11]. Recognized by the National Academies as a new science [20], its key areas of research have been the study of information maneuvers, motive identification, and information diffusion, as well as the evaluation of the effectiveness of information campaigns and mitigation strategies. Though the field has most extensively focused on the spread of political disinformation,

it has more recently expanded to tackle the problem of medical misinformation [11].

Of particular concern in social cybersecurity is the existence of automated accounts on social media platforms since they are used to spread online disinformation and influence elections [21]. They have also manipulated public health discourse by propagating misinformation on topics such as e-cigarettes, diets, and medications [22]. Because of the influence of these bots on public opinion, several studies have been conducted on the use of bots for spreading vaccine information [23,24]. Before the COVID-19 pandemic, Broniatowski et al [25] examined the extent to which bots spread anti-vaccine messages, showing the high rates of vaccine content they spread and comparing it with the effects of Russian trolls, whose messages primarily sought to increase discord online. Dyer [26] determined that after Russian trolls, bots were the most prolific vaccine-related tweeters. Huang and Carley [27] found that accounts linking to coronavirus information from less reliable sites were more likely to be bots. Ng and Carley [28] also found that bots change vaccine stance more easily than non-bots. Hence, understanding the actions of automated accounts is a crucial part of vaccine-related online influence campaigns.

The BEND Framework

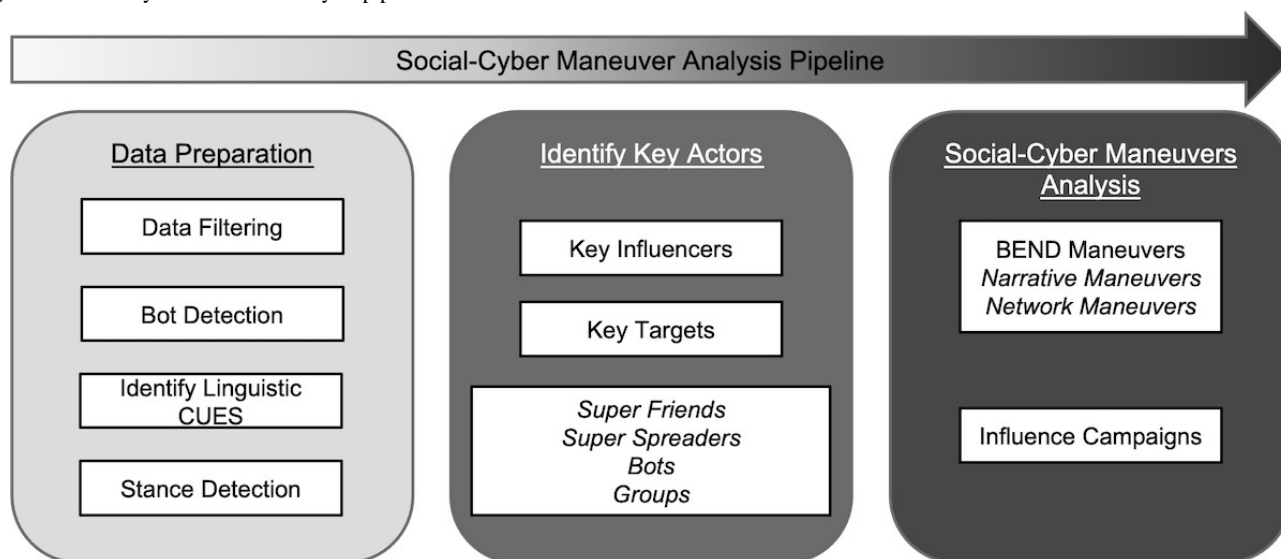
A crucial component of social cybersecurity's efforts to characterize online influence operations has been the struggle to establish the motives and tactics of those seeking to manipulate conversations in cyberspace. The BEND framework was developed to assist in the theoretical conceptualization of this problem by providing a taxonomy of 16 categories of maneuvers for conducting online influence [29]. These categories are divided into 2 types: narrative and network maneuvers. These types are further divided into positive and negative directions of influence. Narrative maneuvers focus on the information and content of messages. These maneuvers affect what is being discussed and how it is discussed. Network maneuvers focus on how the network and communities are shaped and the positions of key actors.

The BEND framework provides analysts and researchers with a way to conceptualize the tactics used in online information operations. Though this framework was discussed in reference to election manipulation [29], it has not yet been applied to vaccine-related influence campaigns.

Methods

In this work, we used a methodology similar to the pipelines in other social cyber-security studies [30]. For the social-cyber maneuver analysis, the end state is to gain a comprehensive understanding of the actors and their maneuvers used to manipulate others on social networks. The pipeline is broken into 3 parts: data preparation, key actor identification, and social-cyber maneuvers analysis (Figure 1).

Figure 1. Social-cyber maneuver analysis pipeline.



Data Preparation

Data Collection

The data used in this work are a subset of COVID-19 tweets collected from Twitter using the Twitter application

Table 1. Keywords used to collect COVID-19 vaccine-related tweets.

Filter	Keywords
Filter 1: COVID-19 tweets	coronavirus, coronavirus, wuhan virus, wuhanvirus, 2019nCoV, NCoV, NCoV2019, covid-19, covid19, covid 19
Filter 2: vaccine tweets	vaccine, vax, mRNA, autoimmuneencephalitis, vaccination, getvaccinated, covidisjustacold, autism, covidshotcount, dose1, dose2, VAERS, GBS, believemothers, mybodymychoice, thisisourshot, killthevirus, proscience, immunization, gotmyshot, igotheshot, covidvaccinated, beatcovid19, moderna, astrazeneca, pfizer, johnson & johnson, j&j, johnson and johnson, jandj

We divided the data into 3 time periods surrounding the introduction of the Pfizer vaccine: December 1-7, 2020 (the week before the rollout), December 8-10, 2020 (during the week of the rollout in the United States and the United Kingdom), and January 25-31, 2021 (6 weeks after the rollout). The 3 periods consisted of 471,962, 694,200, and 662,776 users and 935,709, 1,511,344, and 1,368,035 tweets, respectively.

Identifying Bots

Using the Tier-1 BotHunter algorithm by Beskow and Carley [31,32], we determined the probability that each user within the data set was a bot. BotHunter is a random forest regression model trained on labeled Twitter data sets. It was developed from forensic analyses of events with extensively reported bot activity, such as the attack against the Atlantic Council Digital Forensic Research Lab in 2017. This machine learning model considers network-level features (such as the number of followers and friends), user-level attributes (including screen name length and account age), and tweet-level features (such as timing and content). For this work, any score of 75% or greater was labeled as a bot to reduce the chance of false positives and ensure that the accounts classified as bots were truly bots (at the expense of missing some bots) [33].

programming interface (API) and keywords related to COVID-19. The data set was then further filtered using the vaccine-related terms shown in Table 1. Furthermore, we removed tweets from non-English speaking users from the data.

Linguistic Cues

The NetMapper software [34] was used to extract linguistic cues from the tweet text. These are metrics helpful in identifying a tweet's sentiment and author's emotional state [11]. Examples of these cues include the frequency of positive and negative terms, types of pronouns, emojis, and others. These tweet attributes are used to identify BEND maneuvers and actors participating in such maneuvers.

Organization Risk Analyzer - PRO Software

The Organization Risk Analyzer (ORA)-PRO software [34] is a dynamic meta-network analysis tool used extensively in this study to examine and characterize key actors, conversations, and the overall structure of the Twitter data. Key features used included a network data visualization tool, stance detection function, Twitter analysis report, and the BEND and Community Assessment report.

Stance Detection

We used the stance detector [35] built into ORA-PRO to divide the data set into the pro-vaccine and anti-vaccine communities. This stance detector starts with a set of hashtags that the user initially labels as pro- and anti- with respect to an issue. The stance detector uses these hashtags to label the stance of the

Twitter accounts that used them. The algorithm then uses the concept of influence propagation to label the stance of users who did not use any of the pre-labeled hashtags. This propagation through the user communication network proceeds by repeating 2 steps. First, users with a known stance are used to determine the stances of some of the hashtags that have not yet been labeled. In this step, hashtags that are used overwhelmingly by users of one stance over the other are accordingly assigned that stance. In the second step, hashtags with a known stance are used to determine the stances of some of the unlabeled users. Users who have overwhelmingly used hashtags of one stance rather than the other are labeled with that stance. Additionally, both steps allow stance to spread directly from user to user. In both steps, unlabeled users who

are predominantly connected to users of the same stance are assigned that stance.

The algorithm also provides a confidence level for each stance classification. After running the stance detector on our data, pro-vaccine users had a mean confidence level of approximately 99% to 100% for each time period. However, the anti-vaccine users for the before, during, and after rollout periods had mean confidence levels of 84%, 85%, and 67%, respectively. [Table 2](#) shows the number of users classified by stance for each time period and the number of tweets by these communities. There were noticeably fewer anti-vaccine users and tweets than pro-vaccine users and tweets. Though the stance detector also identified neutral nodes, we excluded them from this study.

Table 2. The number of users labeled as pro-vaccine and anti-vaccine, along with the number of tweets by users of each stance after running the stance detector.

Time period	Users labeled by stance detection		Number of tweets by users of each stance	
	Pro-vaccine	Anti-vaccine	Pro-vaccine	Anti-vaccine
Before rollout	216,156	36,609	186,726	31,200
During rollout	195,334	47,566	292,607	55,406
After rollout	430,278	19,519	338,035	30,560

Examining Key Actors and Social-Cyber Maneuvers

We used reports within ORA-PRO to gain insight on key actors, individual tweets, BEND maneuvers [36], and the entire network. We ran the reports and analyzed each of the 3 time periods on each of the subsequent stance communities. By examining an entire time period, we observed the interactions of the users between those of opposing or neutral stances. Then, by focusing on the individual communities by stance, we conducted a more fine-grained analysis.

ORA-PRO's Twitter report allows us to identify and analyze key agents or actors, hashtags, tweets, and other Twitter attributes on Twitter data. Key actors are useful in understanding who are the most influential entities and what are the most influential conversations. The first type of key actors we observed was super friends, or users that exhibit frequent 2-way communication with others, such as reciprocal mentioning or retweeting. The second type that we examined was super spreaders. These users generate content that is shared often, facilitating the diffusion of information across the network. We then extracted the list of tweets and hashtags for each of these key actors for further inspection.

Additionally, the Twitter report identified valuable tweets. In this study, we specifically focused on the most propagated tweets within a data set. These are tweets that have the highest combined values for retweets, replies, and quotes. This information aided in understanding social-cyber maneuver narratives and actions.

The ORA-PRO software uses NetMapper's linguistic cues as input for detecting BEND maneuvers in tweets using the BEND and Community Assessment report. Of the most propagated tweets, we used this report in conjunction with manual labeling to gain insight into the social-cyber maneuvers used within our data sets.

In our analysis, we organized the BEND maneuvers into 1 of 6 application categories based on the similarity of the maneuvers: developing the narrative, emotional influence, countering the narrative, affecting leaders, making or growing groups, or dissolving or reducing groups ([Table 3](#)). These represent macro-level actions occurring as a result of multiple BEND maneuvers. We observed these combinations over time to identify a concerted effort to influence target audiences of their stances. We labeled the narratives and actions for the 100 most propagated tweets for each stance community within each time period and grouped them into these application categories.

Table 3. BEND maneuvers organized into application categories.

BEND maneuver and application categories	Maneuvers
Narrative maneuvers	
Developing narrative	Engage, explain, enhance
Emotional influence	Excite, dismay
Countering narrative	Distract, dismiss, distort
Network maneuvers	
Affecting leaders	Back, neutralize
Making groups	Build, boost, bridge
Reducing groups	Neglect, narrow, nuke

Results

Key Actor Overview Overtime

Key Influencers: Super Friends and Super Spreaders

We used ORA-PRO to calculate the super friends and super spreaders for the 3 data sets for each time period and identified the types of entities that fell into each of these categories.

The top 10 super friends throughout the 3 data sets were predominately pro-vaccine, though varied in the types of actors. All of the top 10 super friends identified before the rollout were unverified Twitter accounts and relatively low-profile users. ORA classified all of the tweets as pro-vaccine, and 3 of them, we identified as amplifier bots [37]. During the rollout, the top 10 included several anti-vaccine users and a single neutral stance user. Of the 3 bots on the list during this period, 2 news bots emerged alongside one of the pro-vaccine bots from the before period. At 6 weeks later, several higher-profile users from health and government organizations appeared as super friends. These

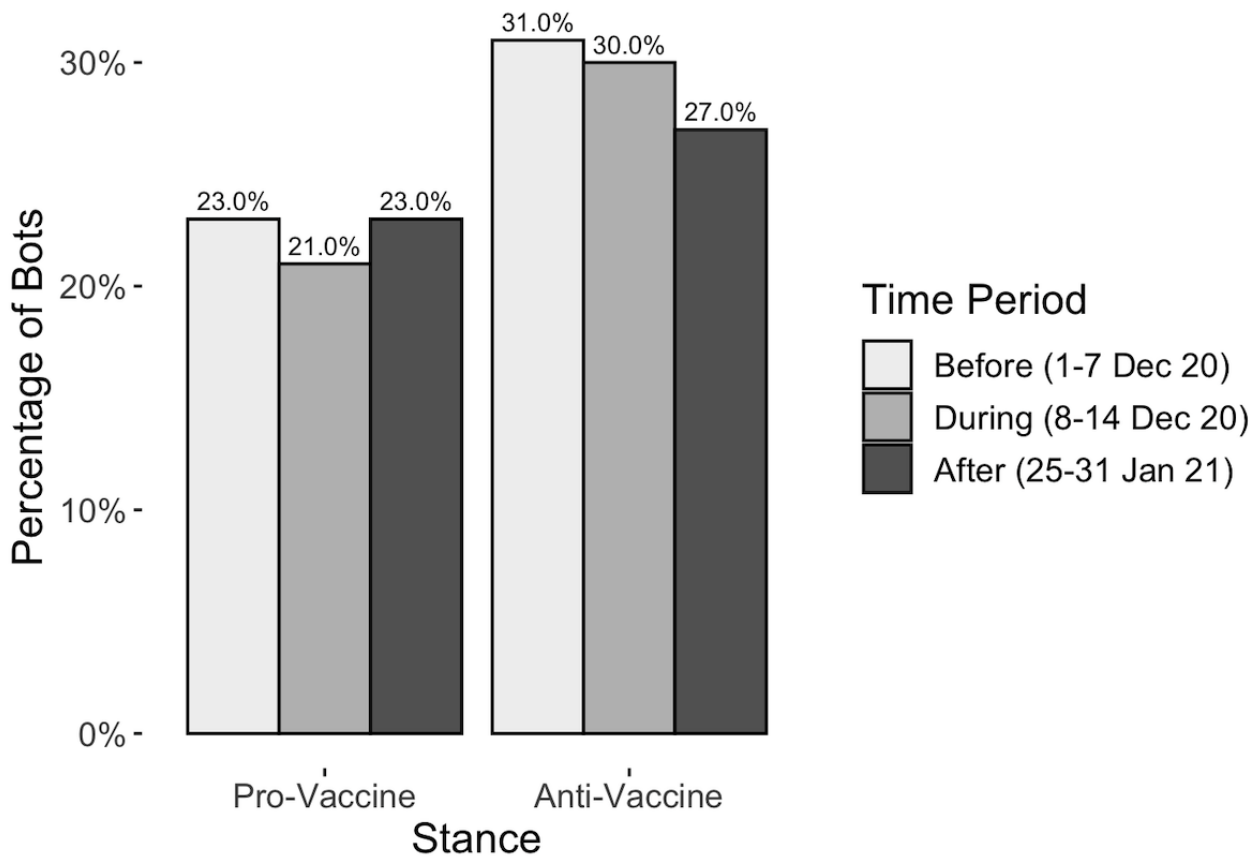
included the World Health Organization, the India Ministry of Health, and the India Official COVID Response account.

Except for one instance, the top 10 super spreaders were either classified as pro-vaccine or neutral within the 3 data sets. All of the users before and during the rollout were high-profile verified Twitter accounts. During these 2 periods, the super spreaders were primarily health organizations, vaccine manufacturers, news organizations, and senior government leaders. After the rollout, the types of accounts identified as super spreaders changed. Though a couple of news organizations and health-related accounts remained on the list, the users were more community leaders or professionals with a substantial reach, such as actors or journalists.

Bot Influencers

The BotHunter results revealed that anti-vaccine agents consisted of a higher percentage of bots than the pro-vaccine agents (Figure 2). Though anti-vaccine bots decreased over time, the number of bots remained relatively higher than the total percentage of pro-vaccine bots of the same time periods.

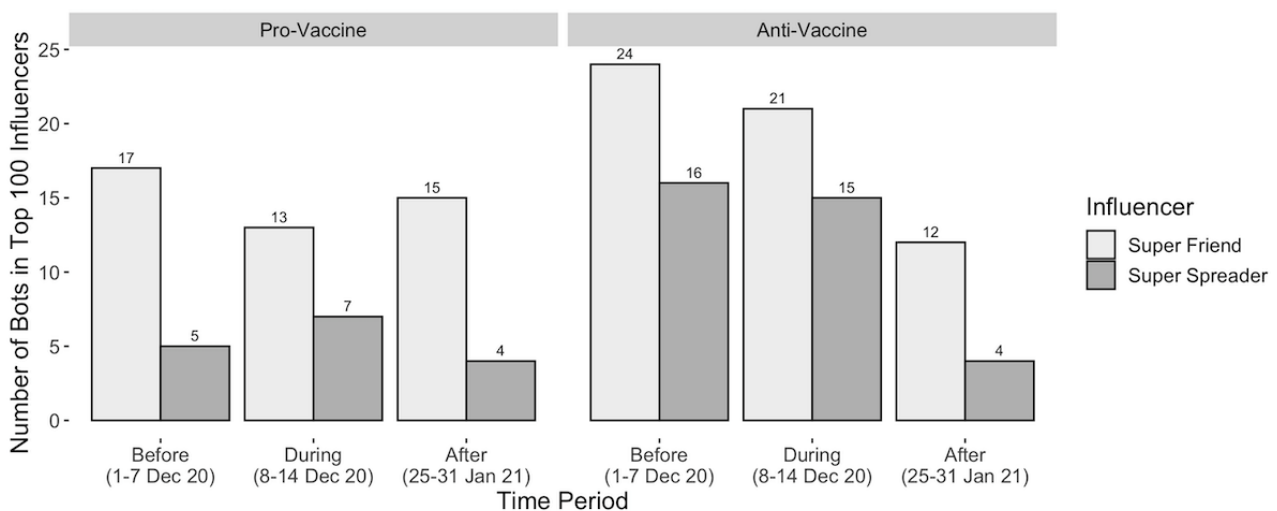
Figure 2. Percentage of bots by stance by time period.



We calculated the number of bots within the top 100 super spreaders and super friends (Figure 3). The high percentage of super friends shows that users are interacting with the bots and engaging in 2-way communications. The super spreaders show that bots are effectively diffusing tweets through the network. These bots have managed to connect with users on Twitter, which makes them susceptible to the information or

disinformation that these bots can be posting. Furthermore, the data show a noticeable decline in the number of anti-vaccine influencers after the rollout. Super spreaders, for example, reduced from 15 and 16 bots during the first 2 periods to only 4 after the rollout. This difference is likely a result of the Twitter policy against anti-vaccine disinformation enacted mid-December 2020 [9].

Figure 3. Number of bots in top 100 influencers.



BEND—Narrative Maneuvers

Using ORA-PRO, we identified the most propagated messages for each of the stances and periods. We then identified and

evaluated how narratives manifested themselves as maneuvers to persuade others. This section provides a cumulation of narrative BEND maneuvers observed within each of the communities.

Pro-Vaccine Communities—Narrative Maneuvers

Pro-vaccine communities had varying messages before the vaccine rollout. Many of them were *excite* messages, defined as messages that elicit a positive emotion such as joy or excitement. Users posted positive tweets about the vaccine’s approval and then encouraged others to get their vaccine when it became available. At the same time, many users also attempted to compel others to curb the growing number of COVID-19–related illnesses and deaths using the *dismay* maneuver. This is messaging to elicit negative emotion such as sadness or anger, to warn users of the consequences of not getting the vaccine. Health officials and organizations used the maneuver, *explain*, which is to educate on a topic using details and relevant facts, to inform the science behind the vaccine and build confidence in its use. To counter vaccine myths, users used the *dismiss* maneuver, the maneuver used to downplay anti-vaccine information as either irrelevant, inconsequential, or foolish. Many users also *dismissed* many of the fears that stemmed from the vaccine’s rapid development. These maneuvers were typically followed up with *explain* maneuvers that attempted to debunk these myths using scientific evidence or detailed forms of justification. Finally, users would *enhance* their pro-vaccine ideas or encourage their views with the support of prominent actors or interesting content. Many, for example, tweeted and quoted articles about 3 former US Presidents volunteering to get the vaccine to promote trust in the vaccine.

During the rollout, pro-vaccine communities continued to post similar types of messages as in the week prior. Many users expressed *excitement* about the first person to receive the Pfizer vaccine, a 90-year-old woman from the United Kingdom leading the vaccine rollout. Other types of optimistic *excite* messages included those from users of various countries approving and purchasing vaccines as well as many *excited* at the sight of the logistics vehicles containing the vaccines within the distribution process. Additionally, pro-vaccine proponents also added to their *explain* messages about how the vaccines work by emphasizing the vaccine’s effectiveness after the first dose and

supporting the overall narrative with charts and results from the vaccine trials during the development process. Furthermore, medical professionals made efforts to *engage* with their more hesitant audiences to instill confidence in the vaccine and encourage them to get vaccinated.

After the rollout, messages continued to *explain* science-backed research for the development, safety, and efficacy of the vaccine to build trust in its use while countering anti-vaccine myths and narratives. Pro-vaccine users during this period showed general *excitement* and optimism about how the vaccine will benefit themselves, their families, and their communities. There were general *excite* tweets about the authorizations and distributions of the vaccine worldwide. Individuals also spread *excite* messages about finally getting the vaccine, getting an appointment for the vaccine, or just desiring to get the vaccine. Many users combined these messages with the *engage* maneuver by taking ownership of the vaccination process by setting the example as a vaccinated individual and encouraging others to also get vaccinated.

Throughout these periods, pro-vaccine communities engaged in hashtag hijacking by tying pro-vaccine narratives to hashtags intuitively associated with anti-vaccine messages. By adding #antivax, #antivaxxer, and other similar anti-vaccine-related hashtags to their tweets, pro-vaccine communities used these hashtags in large numbers to draw attention to pro-vaccine messages with anti-vaccine keywords (Table 4). In one case, they used it to *enhance* the pro-vaccine messages, typically by attaching this hashtag to pro-vaccine *explain* messages intended for vaccine hesitant users. In another case, hashtag hijacking tied the hashtag to satirical messages related to anti-vaccine individuals’ actions. The pro-vaccine message *distorts* the anti-vaccine message with a quote or a reply or somehow ties their narrative to a specific anti-vaccine incident. Furthermore, in some uses of the hashtags, pro-vaccine users *engaged* anti-vaccine users to condemn or insult them for either spreading disinformation or other anti-vaccine behavior.

Table 4. Hashtag hijacking: usage count of anti-vaccine–related hashtags by pro-vaccine users.

Hashtag	Before rollout (n=2218), n	During rollout (n=1221), n	After rollout (n=768), n
antivaccination	0	26	5
antivaccine	55	47	68
antivax	457	281	247
antivaxer	5	3	0
antivaxers	26	11	13
antivaxx	83	54	63
antivaxxer	133	39	62
antivaxxers	1459	760	310

Anti-Vaccine Communities—Narrative Maneuvers

In the week leading up to the vaccine approval and distribution, users were already expressing their COVID-19 anti-vaccine views on social media. The most popular types of messages were the emotionally appealing *dismay* messages about side

effects from the vaccine. Anti-vaccine users shared messages about how the vaccine causes female infertility, destroys the immune system, or leads to death. These messages were further *enhanced* with references to scientists, doctors, former Pfizer representatives, and politicians. In many of these messages, the side effects were *explained* using plausible arguments and

pseudoscientific methods and information. To counter pro-vaccine messages, anti-vaccine proponents attempted tactics such as *dismissing* the vaccine effectiveness, suggesting that a person's immune system is more than sufficient against the virus. They also countered with the *distract* maneuver, which uses misdirection by making other topics seem more important. In one example, the 3 US Presidents volunteering for the vaccine mentioned earlier sought to build confidence. However, opponents of the vaccine focused on *distracting* their audiences with negative political news from the Presidents' pasts relations with China. These insinuated negative links between the Presidents and the country where the virus first began to spread. In another narrative, messages specifically targeting pro-life supporters described the use of fetal cells derived from an abortion during the vaccine development process. These began as *dismay* messages to anger pro-life supporters about its use and then was supported with *explanations* on how the different vaccines used the fetal cells in different phases of the process, making some vaccines more ethical than the others. Proponents then *enhanced* these messages by attaching supportive messages from major religious organizations.

During the rollout, anti-vaccine narratives continued to emphasize many of the negative aspects of the vaccine. Still, many *dismaying* messages about the vaccine side effects dominated anti-vaccine conversations. New *explaining* messages to support these *dismaying* claims included citing the vaccines' published lists of adverse effects and a cost-benefit analysis on the benefits versus the severe reactions resulting from getting the vaccine. Again, these were *enhanced* and validated with statements from medical professionals and scientists. Additional *dismaying* messages emerged as popular during this early period. Topics included news reports for the vaccine causing false positives for HIV, government cautions for allergic reactions to the vaccines, claims that the vaccine is not Halal certified under Islamic dietary laws, and negative experiences from those who participated in the vaccine trials. *Distort* messages, or discussion that alters the main message, helped anti-vaccine messages counter many positive pro-vaccine narratives and propagate anti-vaccine conspiracy theories. They countered the scientific facts about the construct of the vaccine with the lack of peer-reviewed literature to support it, spread manipulated images of Dolly Parton purporting that the vaccine caused her to have Bell's palsy, and suggested that the mRNA vaccines contain nanobots and can change a person's DNA. Furthermore, general anti-vaccine messages from both medical and nonmedical users within this community *engaged* online to express their distrust with the vaccine and recommend not getting the shot without knowing the long-term safety data.

After the rollout, the decrease in anti-vaccine users resulted in spreading fewer anti-vaccine messages. By this time, Twitter removed many anti-vaccine users and their messages for violating their policy on spreading false or misleading COVID-19 vaccine information. The messages that remained, however, were still primarily *dismaying* and *distorting* messages about the adverse side effects and deaths resulting from vaccinations. Despite the Twitter policy, several *distorting* conspiracy theories such as vaccines connecting one's body to cryptocurrency and altering DNA still appeared in the data set.

Furthermore, users continued to *engage* their audiences more practically by expressing hesitancy for a quickly developed vaccine without data on its long-term effects.

BEND—Network Maneuvers

Using the messages, we identified instances of the communities engaging in network maneuvers. Network maneuvers alter the structure of the network by encouraging connections or disconnections between users. In Twitter, one effective tool and indication of a network maneuver is the use of mentions. These types of maneuvers, however, can exist without them.

There were several ways that pro-vaccine communities engaged in network maneuvers. The most common maneuvers were *building* and *boosting*, used for creating a group or to grow the size of a group, respectively. The primary goal for pro-vaccine communities was to urge others to get the vaccine under the premise that the more people who supported and received the vaccine, the sooner the pandemic would end. Simultaneously, these groups engaged in the counter maneuver of *narrowing* and *neglecting* to reduce the size of or marginalize the opposing anti-vaccine community. One of the most effective actions for group reduction was using the *nuke* maneuver to dismantle or show the appearance of a dismantled anti-vaccine community. Twitter attempted this maneuver when it created its policy against COVID-19 vaccine disinformation, affecting the entire after-rollout period data set. Another common network maneuver was *backing*, which is an action that increases the importance of leaders or creates new leaders. Pro-vaccine users showed support for government officials, leaders in the medical field, health organizations, and vaccine manufacturers with positive messages and references to these leaders or organizations.

The anti-vaccine community conducted similar network maneuvers to the pro-vaccine community. They aimed to *build* and *boost* their group and reduce the pro-vaccine community using *narrowing* and *neglecting*. This community, however, did not have as many leaders as its opponents. The few that they *backed* included critics of pro-vaccine policies such as an ex-Pfizer vice president, politicians, and scientists who petitioned against the vaccine for safety concerns. Anti-vaccine users, however, had a large selection of leaders that they attempted to *neutralize* or decrease in importance. These opposing leaders were largely the same people and organizations the pro-vaccine community *backed*.

Social-Cyber Maneuvers Applications Over Time

The different narratives and BEND maneuvers from each stance community were associated with one or more application categories. Though each community used different content for their messaging, they used roughly the same techniques. Many of these techniques are used in combination over time to develop more impactful influence campaigns.

Over time, pro-vaccine communities consistently used mostly positive narrative content in their messaging while applying a pattern of developing their narrative, using emotional influence, and countering the opposing community's narrative (Figure 4). *Excite* messages combined with the narratives about the approval, distribution, and administration of the vaccine were

utilized bots to a greater extent to *build* their communities and spread their narratives than did the pro-vaccine community, effectively positioning them as key influencers among anti-vaccine users.

Pro-vaccine users repeated many of the same maneuvers throughout each time period, varying in different narratives that emerged as the rollout occurred. Many of the maneuvers were positive or growth-type maneuvers. They developed narratives around science-based facts *explaining* the safety and effectiveness of the vaccine and emotionally influencing narratives that *excited* their audiences about the health and societal benefits of everyone receiving the vaccine and, to a lesser extent, *dismayed* them with the fatal consequences of not vaccinating. Many of the topics used to develop these narratives were also used to counter anti-vaccine messages. As many anti-vaccine conversations revolved around side effects and conspiracy theories, pro-vaccine proponents rebutted with facts to *explain* the errors in these messages. These narratives were used consistently by highly connected leaders within the community and actors that maintained a high profile apart from Twitter, such as government leaders, news organizations, and medical professionals. Pro-vaccine users tended to *back* government officials and health organizations with positive messaging to build confidence in the proponents of the vaccine as well as the vaccine itself.

Throughout each of the time periods, the anti-vaccine community used similar maneuvers to those the pro-vaccine community used but with a greater frequency of the negative or reducing-type maneuvers. They developed focused narratives and countered pro-vaccine messaging to increase hesitancy and doubt. Their most consistent technique was using *dismaying* messages of the adverse side effects, the uncertainty of the long-term effects, and vaccine deaths. Conspiracy theories about the vaccine also added to the anti-vaccine narrative. Users attempted to *neutralize* pro-vaccine leaders by discrediting them and their associated messaging, and they *backed* leaders who criticized the vaccine and encouraged others not to get the vaccine.

Finally, as the host for the pro-vaccine and anti-vaccine engagements, Twitter is in a unique position with the ability to filter the discussion on their platform. Their policy to remove misleading and false anti-vaccine *nuked* the anti-vaccine community by significantly reducing the number of anti-vaccine users and tweets that had grown at the time of the rollout. The social media site made a policy to fight disinformation, which resulted in supporting the pro-vaccine effort to reduce the size of the anti-vaccine community and messaging.

Limitations

One major limitation was the ability for the stance detection to separate the nodes into pro-vaccine and anti-vaccine communities. First, many pro-vaccine maneuvers use neutral hashtags. Second, anti-vaccine mean confidence levels for each

time period were lower than those for pro-vaccine, with the after-rollout data only reaching as high as 67%, even after multiple iterations of hashtag labeling. Third, hashtag latching made stance selection difficult as some hashtags commonly used for pro- or anti-vaccine messages were used to gain the attention of members of the other community. Therefore, further study is required for improving the separation between pro- and anti-vaccine agents and tweets.

A second limitation is the ability of ORA-PRO to detect BEND maneuvers. This required manual verification of select entities within the data set. Newer versions of ORA-PRO continue to refine the metrics to better identify some of the maneuvers, especially network maneuvers, that occur over time. The results of this study inform the specifications and thresholds for improving the software.

Finally, many tweets and users that existed during the initial data collection were either deleted or suspended due to Twitter Rules violations. This made it difficult for observing historical tweets with their associated images and videos as well as visualizing tweets with replies and mentions within the Twitter environment.

Conclusions

In this paper, we analyzed how pro-vaccine and anti-vaccine communities around the initial COVID-19 vaccine administration attempted to persuade others of their stance under the BEND maneuvers framework. The BEND maneuvers allowed us to examine the different techniques used by each community. We observed the actions of different types of key actors within the different groups and analyzed their varying techniques. Additionally, we examined the main concepts and messages of tweets tweeted within each community and the extent they acted as each of these social-cyber maneuvers. These maneuvers were combined into application categories to gain a macro-level understanding of how the maneuvers were used in combination as an overarching influence campaign over time.

Real world events influence online discussions, and over time, the changes in these conversations reflect changes in beliefs. In this case, the efforts of these 2 communities can lead users to either vaccinate themselves against COVID-19 or not, possibly changing the direction of the pandemic. Future work should look at how these changes in beliefs mobilize into changes in behavior. Furthermore, though many influencing actions result from users interacting with other users, the policies that govern the use of social media can impact the size of a community and their ability to spread their narrative throughout the network. Therefore, research to regularly detect and evaluate the effectiveness of social-cyber maneuvers and make pointed network structure alterations based on specific narratives is needed to understand the consequences of different interventions and implement better policies to impact influence campaigns on social media.

Acknowledgments

The research for this paper was supported in part by the Office of Naval Research (ONR) under grant N00014182106, the Knight Foundation, the United States Army, and the center for Informed Democracy and Social-cybersecurity (IDeaS). The views and conclusions are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Knight Foundation, the ONR, the United States Army, or the US Government.

Conflicts of Interest

KMC is the author of the BEND software, which is now available in the commercial tool, Organization Risk Analyzer - PRO (ORA-PRO).

References

1. COVID-19 Dashboard. Johns Hopkins Coronavirus Research Center. URL: <https://coronavirus.jhu.edu/map.html> [accessed 2022-02-16]
2. Bonnevie E, Gallegos-Jeffrey A, Goldberg J, Byrd B, Smyser J. Quantifying the rise of vaccine opposition on Twitter during the COVID-19 pandemic. *Journal of Communication in Healthcare* 2020 Dec 15;14(1):12-19. [doi: [10.1080/17538068.2020.1858222](https://doi.org/10.1080/17538068.2020.1858222)]
3. Porter D, Porter R. The politics of prevention: anti-vaccinationism and public health in nineteenth-century England. *Med Hist* 1988 Jul 16;32(3):231-252 [FREE Full text] [doi: [10.1017/s0025727300048225](https://doi.org/10.1017/s0025727300048225)] [Medline: [3063903](https://pubmed.ncbi.nlm.nih.gov/3063903/)]
4. Poland GA, Spier R. Fear, misinformation, and innumerates: how the Wakefield paper, the press, and advocacy groups damaged the public health. *Vaccine* 2010 Mar 11;28(12):2361-2362. [doi: [10.1016/j.vaccine.2010.02.052](https://doi.org/10.1016/j.vaccine.2010.02.052)] [Medline: [20199766](https://pubmed.ncbi.nlm.nih.gov/20199766/)]
5. Dubé E, Laberge C, Guay M, Bramadat P, Roy R, Bettinger JA. Vaccine hesitancy: an overview. *Hum Vaccin Immunother* 2013 Aug 27;9(8):1763-1773 [FREE Full text] [doi: [10.4161/hv.24657](https://doi.org/10.4161/hv.24657)] [Medline: [23584253](https://pubmed.ncbi.nlm.nih.gov/23584253/)]
6. Shahsavari S, Holur P, Wang T, Tangherlini T, Roychowdhury V. Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news. *J Comput Soc Sci* 2020;1 [FREE Full text] [doi: [10.21203/rs.3.rs-52079/v1](https://doi.org/10.21203/rs.3.rs-52079/v1)]
7. Cornwall W. Officials gird for a war on vaccine misinformation. *Science* 2020 Jul 03;369(6499):14-15. [doi: [10.1126/science.369.6499.14](https://doi.org/10.1126/science.369.6499.14)] [Medline: [32631873](https://pubmed.ncbi.nlm.nih.gov/32631873/)]
8. Jamison AM, Broniatowski DA, Dredze M, Sangraula A, Smith MC, Quinn SC. Not just conspiracy theories: Vaccine opponents and proponents add to the COVID-19 'infodemic' on Twitter. *Harv Kennedy Sch Misinformation Rev* 2020 Sep 9;1:1 [FREE Full text] [doi: [10.37016/mr-2020-38](https://doi.org/10.37016/mr-2020-38)] [Medline: [34368805](https://pubmed.ncbi.nlm.nih.gov/34368805/)]
9. COVID-19: Our approach to misleading vaccine information. Twitter Blog. URL: https://blog.twitter.com/en_us/topics/company/2020/covid19-vaccine.html [accessed 2022-02-16]
10. Updates to our work on COVID-19 vaccine misinformation. Twitter Blog. URL: https://blog.twitter.com/en_us/topics/company/2021/updates-to-our-work-on-covid-19-vaccine-misinformation.html [accessed 2022-02-16]
11. Carley KM. Social cybersecurity: an emerging science. *Comput Math Organ Theory* 2020 Nov 16;26(4):365-381 [FREE Full text] [doi: [10.1007/s10588-020-09322-9](https://doi.org/10.1007/s10588-020-09322-9)] [Medline: [33223952](https://pubmed.ncbi.nlm.nih.gov/33223952/)]
12. Kummervold PE, Martin S, Dada S, Kilich E, Denny C, Paterson P, et al. Categorizing vaccine confidence with a transformer-based machine learning model: analysis of nuances of vaccine sentiment in Twitter discourse. *JMIR Med Inform* 2021 Oct 08;9(10):e29584 [FREE Full text] [doi: [10.2196/29584](https://doi.org/10.2196/29584)] [Medline: [34623312](https://pubmed.ncbi.nlm.nih.gov/34623312/)]
13. Krishnan G, Sowmya Kamath S, Sugumaran V. Predicting vaccine hesitancy and vaccine sentiment using topic modeling and evolutionary optimization. In: Métais E, Meziane F, Horacek H, Kapetanios E, editors. *Natural Language Processing and Information Systems. NLDB 2021. Lecture Notes in Computer Science*, vol 12801. Cham, Switzerland: Springer; 2021:255-263.
14. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrocioni W. Polarization of the vaccination debate on Facebook. *Vaccine* 2018 Jun 14;36(25):3606-3612. [doi: [10.1016/j.vaccine.2018.05.040](https://doi.org/10.1016/j.vaccine.2018.05.040)] [Medline: [29773322](https://pubmed.ncbi.nlm.nih.gov/29773322/)]
15. Memon SA, Tyagi A, Mortensen DR, Carley KM. Characterizing sociolinguistic variation in the competing vaccination communities. 2020 Presented at: Social, Cultural, and Behavioral Modeling: 13th International Conference, SBP-BRiMS; October 18-21, 2020; Washington, DC. [doi: [10.1007/978-3-030-61255-9_12](https://doi.org/10.1007/978-3-030-61255-9_12)]
16. Bonnevie E, Goldberg J, Gallegos-Jeffrey AK, Rosenberg SD, Wartella E, Smyser J. Content themes and influential voices within vaccine opposition on Twitter, 2019. *Am J Public Health* 2020 Oct;110(S3):S326-S330. [doi: [10.2105/ajph.2020.305901](https://doi.org/10.2105/ajph.2020.305901)]
17. Johnson NF, Velásquez N, Restrepo NJ, Leahy R, Gabriel N, El Oud S, et al. The online competition between pro- and anti-vaccination views. *Nature* 2020 Jun 13;582(7811):230-233. [doi: [10.1038/s41586-020-2281-1](https://doi.org/10.1038/s41586-020-2281-1)] [Medline: [32499650](https://pubmed.ncbi.nlm.nih.gov/32499650/)]
18. Thelwall M, Kousha K, Thelwall S. Covid-19 vaccine hesitancy on English-language Twitter. *EPI* 2021 Mar 15;1. [doi: [10.3145/epi.2021.mar.12](https://doi.org/10.3145/epi.2021.mar.12)]

19. Boucher J, Cornelson K, Benham JL, Fullerton MM, Tang T, Constantinescu C, et al. Analyzing social media to explore the attitudes and behaviors following the announcement of successful COVID-19 vaccine trials: infodemiology study. *JMIR Infodemiology* 2021;1(1):e28800 [FREE Full text] [doi: [10.2196/28800](https://doi.org/10.2196/28800)] [Medline: [34447924](https://pubmed.ncbi.nlm.nih.gov/34447924/)]
20. National Academies of Sciences, Engineering, and Medicine. *A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis*. Washington, DC: The National Academies Press; 2019.
21. 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](https://doi.org/10.1145/2818717)]
22. Allem J, Ferrara E. Could social bots pose a threat to public health? *Am J Public Health* 2018 Aug;108(8):1005-1006. [doi: [10.2105/AJPH.2018.304512](https://doi.org/10.2105/AJPH.2018.304512)] [Medline: [29995482](https://pubmed.ncbi.nlm.nih.gov/29995482/)]
23. 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. [doi: [10.1177/2056305119865465](https://doi.org/10.1177/2056305119865465)]
24. Liang C, Li Y, Luo J. A novel method to detect functional microRNA regulatory modules by bicliques merging. *IEEE/ACM Trans Comput Biol Bioinform* 2016;13(3):549-556. [doi: [10.1109/TCBB.2015.2462370](https://doi.org/10.1109/TCBB.2015.2462370)] [Medline: [27295638](https://pubmed.ncbi.nlm.nih.gov/27295638/)]
25. 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](https://doi.org/10.2105/AJPH.2018.304567)] [Medline: [30138075](https://pubmed.ncbi.nlm.nih.gov/30138075/)]
26. Dyer O. Vaccine safety: Russian bots and trolls stoked online debate, research finds. *BMJ* 2018 Aug 30;362:k3739. [doi: [10.1136/bmj.k3739](https://doi.org/10.1136/bmj.k3739)] [Medline: [30166280](https://pubmed.ncbi.nlm.nih.gov/30166280/)]
27. Huang B, Carley KM. Disinformation and misinformation on Twitter during the novel coronavirus outbreak. Cornell University. 2020. URL: <https://arxiv.org/abs/2006.04278> [accessed 2022-02-16]
28. Ng LHX, Carley K. Flipping stance on bot? and non bot? COVID vaccine stance. Cornell University. 2021 Jun 21. URL: <https://arxiv.org/abs/2106.11076> [accessed 2022-02-16]
29. Beskow DM, Carley KM. Social cybersecurity: an emerging national security requirement. *Military Review* 2019;99(2):1 [FREE Full text]
30. Uyheng J, Magelinski T, Villa-Cox R, Sowa C, Carley KM. Interoperable pipelines for social cyber-security: assessing Twitter information operations during NATO Trident Juncture 2018. *Comput Math Organ Theory* 2019 Sep 28;26(4):465-483. [doi: [10.1007/s10588-019-09298-1](https://doi.org/10.1007/s10588-019-09298-1)]
31. Beskow DM, Carley KM. Bot-hunter: a tiered approach to detecting & characterizing automated activity on Twitter. 2018 Presented at: 11th International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation; July 10-13, 2018; Washington, DC.
32. Beskow DM, Carley KM. Its all in a name: detecting and labeling bots by their name. *Comput Math Organ Theory* 2018 Dec 18;25(1):24-35. [doi: [10.1007/s10588-018-09290-1](https://doi.org/10.1007/s10588-018-09290-1)]
33. Ng LHX, Robertson DC, Carley K. On the stability of BotHunter scores. 2021 Jul 12 Presented at: Center for Informed Democracy & Social Cybersecurity (IDEAS) Conference; July 12-13, 2021; Virtual event URL: <https://www.cmu.edu/ideas-social-cybersecurity/events/conference-archive/archive-conference-2021/conference-papers-2021/conference-2021-paper-17-bothunter.pdf>
34. Carley LR, Reminga J, Carley KM. ORA & NetMapper. 2018 Presented at: 11th International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation; July 10-13, 2018; Washington, DC.
35. Kumar S. Social Media Analytics for Stance Mining A Multi-Modal Approach with Weak Supervision. Carnegie Mellon University. URL: [https://kithub.cmu.edu/articles/thesis/Social Media Analytics for Stance Mining A Multi-Modal Approach with Weak Supervision/12352079](https://kithub.cmu.edu/articles/thesis/Social%20Media%20Analytics%20for%20Stance%20Mining%20A%20Multi-Modal%20Approach%20with%20Weak%20Supervision/12352079) [accessed 2022-02-16]
36. Carley KM. BEND: a framework for social cybersecurity. *Future Force* 2020;6(2):22-27 [FREE Full text]
37. Beskow D. Finding and Characterizing Information Warfare Campaigns. Carnegie Mellon University. 2020. URL: [https://kithub.cmu.edu/articles/thesis/Finding and Characterizing Information Warfare Campaigns/12303089](https://kithub.cmu.edu/articles/thesis/Finding%20and%20Characterizing%20Information%20Warfare%20Campaigns/12303089) [accessed 2022-02-16]

Abbreviations

- API:** application programming interface
- BERT:** Bidirectional Encoder Representations from Transformers
- E/I Index:** external/internal index
- EMA:** European Medicines Agency
- FDA:** Food and Drug Administration
- ORA-PRO:** Organization Risk Analyzer - PRO

Edited by C Basch; submitted 04.10.21; peer-reviewed by M Lotto, J Li, MY Song; comments to author 27.11.21; revised version received 18.12.21; accepted 08.01.22; published 07.03.22.

Please cite as:

Blane JT, Bellutta D, Carley KM

Social-Cyber Maneuvers During the COVID-19 Vaccine Initial Rollout: Content Analysis of Tweets

J Med Internet Res 2022;24(3):e34040

URL: <https://www.jmir.org/2022/3/e34040>

doi: [10.2196/34040](https://doi.org/10.2196/34040)

PMID: [35044302](https://pubmed.ncbi.nlm.nih.gov/35044302/)

©Janice T Blane, Daniele Bellutta, Kathleen M Carley. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 07.03.2022. 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.

Original Paper

Understanding Public Perceptions of Per- and Polyfluoroalkyl Substances: Infodemiology Study of Social Media

Hao Tian¹, PhD; Christy Gaines¹, PhD; Lori Launi², MA; Ana Pomales³, MSc; Germaine Vazquez¹, MSc; Amanda Goharian⁴, BA; Bradley Goodnight⁴, PhD; Erica Haney⁴, PhD; Christopher M Reh², PhD; Rachel D Rogers³, PhD

¹Office of Director, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA, United States

²Office of Associated Director, Agency for Toxic Substances and Disease Registry, Atlanta, GA, United States

³Office of Community Health Hazard Assessment, Agency for Toxic Substances and Disease Registry, Atlanta, GA, United States

⁴Guidehouse LLP, Atlanta, GA, United States

Corresponding Author:

Hao Tian, PhD

Office of Director

National Center for Environmental Health

Centers for Disease Control and Prevention

4770 Buford Highway NE, MS S106-7

Atlanta, GA, 30341

United States

Phone: 1 404 718 5708

Email: ejq7@cdc.gov

Abstract

Background: Per- and polyfluoroalkyl substances (PFAS) are environmental contaminants that have received significant public attention. PFAS are a large group of human-made chemicals that have been used in industry and consumer products worldwide since the 1950s. Human exposure to PFAS is a growing public health concern. Studies suggest that exposure to PFAS may increase the risk of some cancers and have negative health impacts on the endocrine, metabolic, and immune systems. Federal and state health partners are investigating the exposure to and possible health effects associated with PFAS. Government agencies can observe social media discourse on PFAS to better understand public concerns and develop targeted communication and outreach efforts.

Objective: The primary objective of this study is to understand how social media is used to share, disseminate, and engage in public discussions of PFAS-related information in the United States.

Methods: We investigated PFAS-related content across 2 social media platforms between May 1, 2017, and April 30, 2019, to identify how social media is used in the United States to seek and disseminate PFAS-related information. Our key variable of interest was posts that mentioned “PFAS,” “PFOA,” “PFOS,” and their hashtag variations across social media platforms. Additional variables included post type, time, PFAS event, and geographic location. We examined term use and post type differences across platforms. We used descriptive statistics and regression analysis to assess the incidence of PFAS discussions and to identify the date, event, and geographic patterns. We qualitatively analyzed social media content to determine the most prevalent themes discussed on social media platforms.

Results: Our analysis revealed that Twitter had a significantly greater volume of PFAS-related posts compared with Reddit (98,264 vs 3126 posts). PFAS-related social media posts increased by 670% over 2 years, indicating a marked increase in social media users’ interest in and awareness of PFAS. Active engagement varied across platforms, with Reddit posts demonstrating more in-depth discussions compared with passive likes and reposts among Twitter users. Spikes in PFAS discussions were evident and connected to the discovery of contamination events, media coverage, and scientific publications. Thematic analysis revealed that social media users see PFAS as a significant public health concern and seek a trusted source of information about PFAS-related public health efforts.

Conclusions: The analysis identified a prevalent theme—on social media, PFAS are perceived as an immediate public health concern, which demonstrates a growing sense of urgency to understand this emerging contaminant and its potential health impacts. Government agencies can continue using social media research to better understand the changing community sentiment on PFAS

and disseminate targeted information and then use social media as a forum for dispelling misinformation, communicating scientific findings, and providing resources for relevant public health services.

(*J Med Internet Res* 2022;24(3):e25614) doi:[10.2196/25614](https://doi.org/10.2196/25614)

KEYWORDS

PFAS; per- and polyfluoroalkyl substances; social media; public perceptions

Introduction

Background

Per- and polyfluoroalkyl substances (PFAS) are a large class of manufactured chemicals that have been widely produced and used in industry and consumer products such as nonstick cookware, water-repellent clothing, stain-resistant fabrics, carpets, and other items since the mid-20th century [1-3]. The most commonly studied PFAS are perfluorooctanoic acid (PFOA) and perfluorooctane sulfonic acid (PFOS) followed by perfluorohexane sulfonic acid and perfluorononanoic acid. Chemical companies and other industry manufacturers began phasing out the use and manufacture of the PFOA and PFOS variants in the early 2000s [2,3]. However, PFAS persist in the environment, especially in certain foods, water sources, people, and wildlife worldwide owing to their long biological half-lives [4]. Some communities nationwide have been and likely will continue to be exposed to drinking water contaminated with PFAS owing to both new and historical exposure [4-7]. Research examining the impact of PFAS has varied; however, some studies indicate that exposure to certain PFAS may increase the risk of some cancers and have negative health impacts on the endocrine, metabolic, and immune systems [3,5-7]. In recent years, PFAS as environmental contaminants have received significant public attention owing to emerging evidence of widespread PFAS contamination, the development of new state and federal drinking water PFAS guidelines, and a handful of newly released PFAS-focused documentaries and news stories [8-13].

One way to assess and understand the increase in public attention to PFAS is to investigate PFAS-related activity on various social media platforms. Social media is a broad term used to describe web-based platforms that allow individuals as well as representatives of institutions (such as news media, government agencies, nongovernmental organizations, and advocacy groups) to publish content and connect with other users. Widespread adoption of social media apps has fundamentally changed the way the public disseminates and shares information [14]. In the past decade, social media platforms have been used to monitor disease outbreak patterns and communicate information to the public during emergency responses for bubonic plague, swine flu, seasonal influenza, and West Nile virus disease [15-21]. As today's digital activity increases, the use of social media platforms to find and share information about public health issues continues to grow.

Some studies have explored the use of social media to communicate information about emerging health concerns such as the health effects of e-cigarettes [19-22]. A recent study [21] observed that Reddit posts included an abundance of requests looking for healthy alternatives to e-cigarettes and that Twitter

posts tended to focus on information seeking related to regulations and policy debates around e-cigarettes. Variation in information seeking and discussion by platform suggests that some platforms may be more successful than others for government agencies in the reach and delivery of public health information and interventions. Therefore, it is important to analyze data from different social media platforms to gain a better understanding of social media users' perceptions of PFAS.

Objectives

Although we are not aware of any studies investigating PFAS-related social media activity, we believe that observing American social media users' attitude toward PFAS across various platforms has the potential to help government agencies better understand public concerns, prevent and address the spread of misinformation, respond to new PFAS-related incidents, and develop targeted communication and outreach efforts, as demonstrated in other studies targeting different health issues [23,24]. This social media research aims to increase understanding of the public's perception of PFAS and inform how best to reach affected communities with the appropriate health information and environmental health prevention services. The public receives health information from various sources and stakeholders. It is important to understand how perceptions of and dialogue about those sources may affect government agencies' ability to effectively communicate health information. By conducting social media research and analysis of PFAS-related posts, government agencies and other partner organizations may become better equipped to respond to public concerns, questions, and requests.

This study has three objectives: (1) to understand how social media is used to share, disseminate, and engage in public discussions of PFAS-related information in the United States; (2) to identify common themes within PFAS-related mentions across social media platforms; and (3) to identify how social media engagement relates to various news events to better anticipate when and where to target outreach efforts using social media.

Methods

Data Collection

We first selected seven social media platforms—Facebook, Twitter, Reddit, YouTube, Nextdoor, Imgur, and Pinterest—that are likely used by the public to share environmental health-related information and investigated the availability and quality of data from these platforms across a recent 2-year window from May 1, 2017, to April 30, 2019. We focused our analysis on a 2-year window based on 2 primary factors. The first is an increase in media attention to PFAS in late 2017. The second is the dynamic and fluid nature of social media

environments and the internet more broadly. Accordingly, we focused on a recent 24-month period that would enable sufficient time to identify patterns. Owing to the identified challenges in the collection, processing, and comparison of the data associated with these disparate social media platforms, we tailored our data collection to focus on Twitter and Reddit, the 2 platforms that provided access to data and metadata, allowing us to achieve the study objectives. We extracted the data on August 7, 2019, for Twitter (N=98,264) and Reddit (n=3126) through the use of NetBase (NetBase Solutions Inc), a third-party software vendor that facilitated access to the platforms' full-stream application programming interfaces (eg, *firehose* application programming interfaces), which included all publicly available content (ie, not content from private accounts) featuring our identified key terms defined below within our specified time frame. On the basis of the scope of our research questions, we did not collect PFAS-related content posted from countries other than the United States.

Ethics Approval

The data used for this study were based on publicly available information only and, thus, the study was determined to be exempt from the Centers for Disease Control and Prevention Institutional Review Board approval.

Variables of Interest

Our key variable of interest was posts on PFAS or common alternatives, including PFOA, PFOS, and the hashtag derivatives of each term (#PFAS, #PFOA, and #PFOS) across social media platforms. These key term variables were selected based on their more prominent volume and use as identified in an initial exploration of PFAS-related key term frequency during the same time frame on Google Trends (data accessed on June 19, 2019).

Posts were classified as either original posts or engagement posts. Original posts were defined as first-instance comments. Engagement posts were secondary posts in the form of a reply or comment (including retweets on Twitter) to an original post. We acknowledge that these types of posts are different; however, we chose to aggregate at the *engagement* level so that engagement could be compared across platforms. On the basis of the study goals, additional variables of interest included post type, Twitter *likes*, date of post, and PFAS event or geographic location of social media user. Below are the key variables and their associated definitions.

- PFAS post (binary): a post, tweet, or comment on social media platforms that uses one or more of the terms PFAS, PFOA, or PFOS
- Post type (binary): original posts refers to first-time comments. Engagement posts refers to any reply, comment, or retweet that responds to or engages with an existing comment. The engagement ratio is the proportion of engagement posts to original posts. Engagement posts are defined similarly across different social media platforms, although they may be referred to with platform-specific language (eg, a retweet on Twitter)
- Likes (count): specific to Twitter data, likes indicates the number of Twitter users that clicked on like on a tweet

- Date (continuous): the day, month, or year of a tweet, post, or comment
- Geography (categorical): the national jurisdiction location of a PFAS-related post

Although original posts measure message reach, engagement posts measure interactive dialogue among social media users. Engagement ratio is defined as the proportion of engagement posts to original posts, which is an indicator of the degree of engagement by users via replies to, comments on, or retweets of an original post. For example, an engagement ratio of 2 suggests that users were 2 times more likely to reply to, comment on, or retweet an existing post than to generate an original post.

We manually reviewed 3 random samples (n=500) of the data sets to identify keywords or phrases for cleaning irrelevant or misleading data. This manual review was used to implement rule-based removal of similar irrelevant or misleading content throughout both data sets. Irrelevant data accounted for <1% of Twitter (254/98,264, 0.26%) and Reddit (18/3126, 0.58%) data. Examples of removed content included advertisements for various cookware products with marketing language such as *PFOA-free* and irrelevant posts of PFAs, an acronym for *protection from abuse* referencing an unrelated domestic violence abuse order.

Quantitative Analysis

We analyzed and identified unique attributes of public engagement across each platform. We used descriptive statistics, including counts and frequencies of the key quantitative variables, to assess the incidence of PFAS discussions and to identify time, geographic, and entity patterns. We explored PFAS-related posts over time to identify spikes and assess their association with PFAS-related events. We obtained maps of tweets and Reddit posts by geographic area using a data visualization package (ggplot2) in the statistical software R (R Foundation for Statistical Computing) [25].

The purpose of the geographic analysis was to demonstrate which states had the highest number of PFAS-related posts broken down by original and engagement posts, count percentage, and the engagement to original post ratio for each state. Understanding geographic patterns in PFAS-related posts more broadly can help government agencies develop future outreach efforts.

That said, geographic metadata are not always provided by a given social media platform. For this analysis, we identified geographic information for tweets based on state-level geographic data available in an individual user's profile. However, Reddit data did not provide geographic metadata for users' posts. To approximate geographic concentrations of Reddit data, we built a complementary proxy using text analysis of geographic subreddit identification (eg, *r/Michigan*) and user mentions of specific jurisdictions. To complement geographic approximation, we also conducted a scan of post titles that directly referenced a specific geographic location. Although not providing an exact comparison, geographic analysis through these 2 approaches provided insight into potentially underlying differences in PFAS conversations by geography. To account

for population variance across states, we normalized the data using 2010 census data to provide a proportional jurisdiction ratio of per 100,000. Our dissimilar approach to obtaining geographic information, coupled with variance in sample sizes between the 2 platforms, is likely to influence geographical findings and related conclusions.

Qualitative Analysis

We used the constant comparison method based on grounded theory [26,27] to inductively analyze posts and comments, coupled with deductive analysis to identify and analyze major discussion themes on social media platforms. A total of 3 people used Microsoft Excel to individually open code 3 subsets of randomly sampled text (1500 in total; 500/1500, 33.33% per subset) for common themes, cross-validation, and collective refinement of themes. Through an iterative process, some coding categories were collapsed into larger concepts until no further themes emerged in the subsequent analysis, suggesting that we had reached saturation of themes. The team subsequently developed search criteria with a unique set of keywords for each theme based on the initial open coding process (see the *Thematic Analysis* section for example keywords for each theme). Themes were not mutually exclusive and, therefore, each unit of analysis had the potential to be counted across multiple themes.

Mixed Methods Analysis

As a follow-up to the qualitative analysis, we conducted a series of regression analyses to test whether posts containing words from the search criteria established in the qualitative analysis differed from one another and from posts that did not use the search criteria in terms of their level of engagement on the Twitter and Reddit platforms. Search criteria were used to separate posts into thematic groups, and the relationship between these nonexclusive groups was established on two indices measuring user engagement: categorization as a reply to an original post and number of *likes* on Twitter. The data set used for comparison included only social media users of the Twitter and Reddit platforms. The level of analysis was posts and comments, and the unit of analysis was comment or post and not individual users.

The analyses consisted of a logistic regression to examine the effect of the 3 thematic groups on a binary outcome of original posts versus engagement posts. In other words, the logistic regression analysis compared each of the 3 themes to one another in terms of the relative quantity of engagement posts to original posts (ie, level of engagement). In addition, we conducted a Poisson regression analysis to look at the effect of the same predictors on a count outcome representing the number of *likes* that a tweet received on Twitter, a secondary indicator of user engagement. As *likes* is a metric that is unique to Twitter and the sample of Reddit data was much smaller than the Twitter sample, these follow-up mixed methods analyses were limited to the Twitter data. We were unable to conduct the same analyses among Reddit data as we did not have access to upvote or downvote information, a measure used to demonstrate that a user likes or dislikes a post. Furthermore, the smaller number of posts for Reddit provided much lower statistical power than the Twitter data, and the limitation for post length for Twitter and not for Reddit limited our ability to compare the 2 platforms.

Results

Platform Volume, Post Type, and Term Use

As seen in Table 1, there were a total of 101,390 PFAS-related posts on Twitter and Reddit during the 2-year window. Twitter had a significantly greater volume of PFAS-related posts compared with Reddit (98,264 vs 3126).

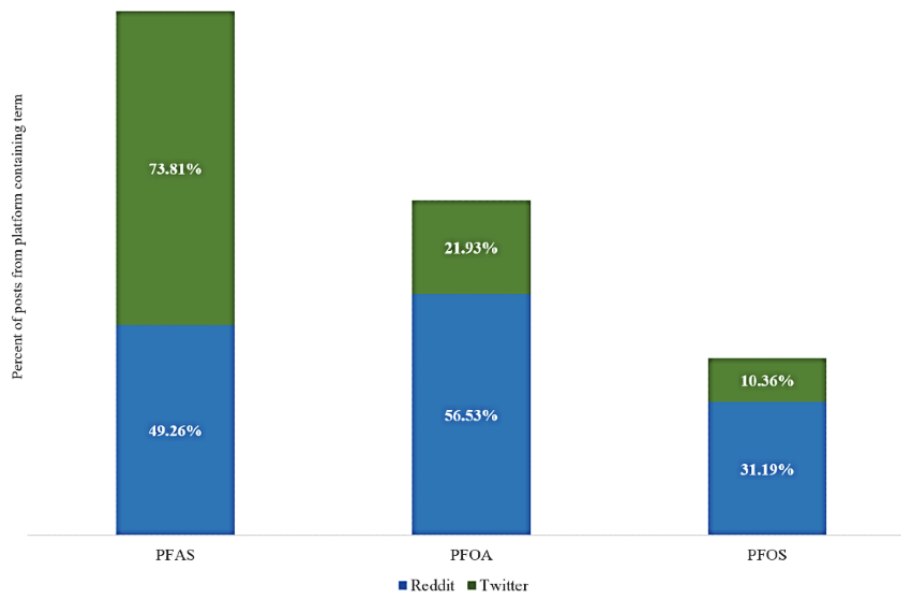
Understanding PFAS-related term use clarifies how social media users were interfacing with PFAS-related conversations on social media. Among Twitter data, social media users used the term *PFAS* in 73.81% (72,528/98,264) of all posts compared with *PFOA* (21,545/98,264, 21.93%) and *PFOS* (10,183/98,264, 10.36%). Given that a post may include multiple terms, the aggregate total resulted in a value of >100%. We saw a different pattern of term use among Reddit posts whereby users most mentioned the term *PFOA* (1767/3126, 56.53%) followed by *PFAS* (1540/3126, 49.26%) and *PFOS* (974/3126, 31.16%; Figure 1).

Table 1. Counts of observations by social media platform.

Platform	Original posts, n (%)	Engagement posts ^a , n (%)	Total, n (%)
Twitter (tweets)	36,795 (37.44)	61,469 (62.55)	98,264 (100)
Reddit (posts)	424 (13.56)	2702 (86.44)	3126 (100)
Total	37,219 (36.71)	64,171 (63.29)	101,390 (100)

^aEngagement posts include all replies and retweets.

Figure 1. Key term distribution across all posts by social media platform (Twitter: N=98,264; Reddit: N=3126). PFAS: per- and polyfluoroalkyl substance; PFOA: perfluorooctanoic acid; PFOS: perfluorooctane sulfonic acid.



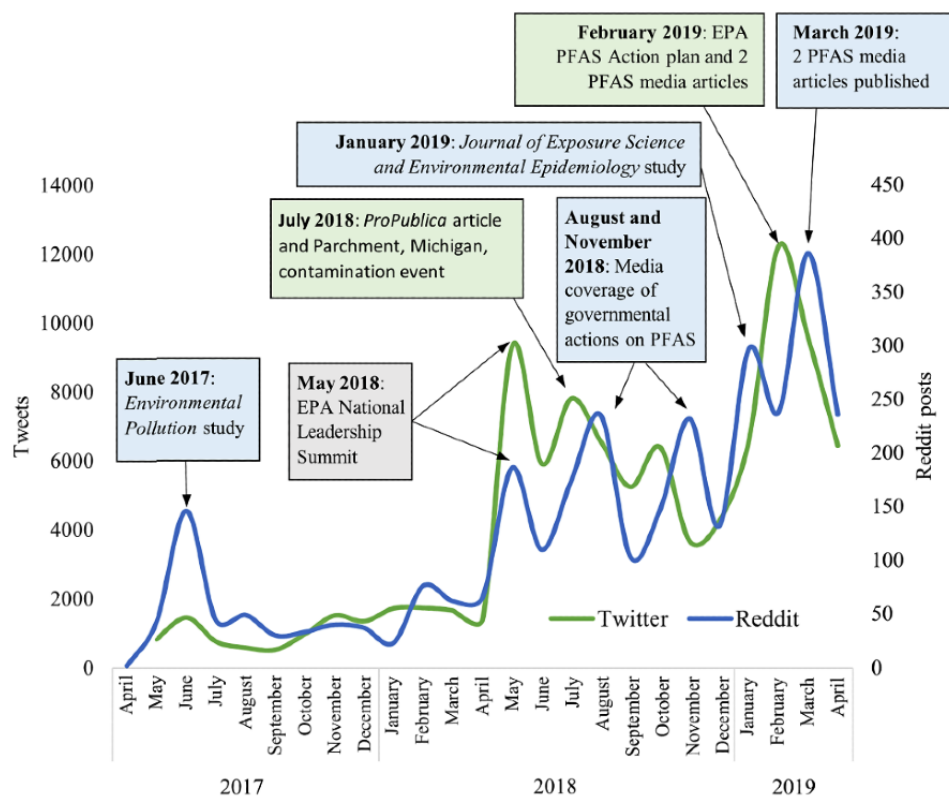
Timing and Event Analysis

Overview

The volume of posts over time underscores a collective rise in interest in and awareness of PFAS. Driving this increase were

9 notable spikes in the volume of public discussion of PFAS (3 on Twitter and 6 on Reddit). We evaluated the data retrospectively and conducted text analysis to determine the primary PFAS-related topics and events mentioned in the posts associated with the spikes (Figure 2).

Figure 2. Number of per- and polyfluoroalkyl substance (PFAS)-related Twitter and Reddit posts by date with annotated events. EPA: Environmental Protection Agency.



Twitter Spikes and Events

The 3 primary spikes in PFAS-related discussions on Twitter occurred in May 2018, July 2018, and February 2019. The content analysis of posts during these periods revealed that the

spike in May 2018 was attributed to the high volume of posts about the US Environmental Protection Agency (EPA) National Leadership Summit (which happened in May 2018). In July 2018, a smaller but notable spike occurred that was attributed to discussions of a ProPublica article on government responses

to PFAS [28] and a local news article about the PFAS water contamination in Parchment, Michigan [29]. Finally, the spike in February 2019 corresponded with the release of a PFAS action plan by the EPA and greater national-level reporting on PFAS by mainstream publications, such as *The New York Times* articles on PFAS exposure in military families [30] and government response [31]. In several instances, spikes occurred simultaneously or were slightly delayed across platforms, indicating the potential for a sustained discussion over time and across platforms of critical PFAS events.

Reddit Spikes and Events

The 6 primary spikes in PFAS-related discussions on Reddit occurred in June 2017, May 2018, August 2018, November 2018, January 2019, and March 2019. The June 2017 spike was primarily related to the publication and release of a peer-reviewed study in *Environmental Pollution* on PFAS exposure in the mid-Ohio River Valley from 1991 to 2012 [32]. The spike in May 2018 was attributed to the high volume of posts about the US EPA National Leadership Summit. A noticeable spike in Reddit activity in August 2018 was related to the publication of 2 different media articles. One was from *ProPublica* on government response [28], the same article found in the Twitter July 2018 spike. The other article was from *MLive*, a local Michigan news site, focused on governmental action toward PFAS contamination [5]. The reprinting of the *ProPublica* article [28] by CNBC on November 12, 2018, led to an increase in Reddit activity, with numerous users posting links to and comments about the article within the subreddit *r/news*. Another noteworthy increase in PFAS-related Reddit activity was largely in response to a peer-reviewed study published in January 2019 in the *Journal of Exposure Science*

and *Environmental Epidemiology* on serum concentrations of PFAS and exposure-related behaviors in African American and non-Hispanic White women [33]. Finally, the March 2019 spike was associated with 2 media articles on the regulatory process of the EPA and PFAS exposure in military families from *The Guardian* [34] and *The New York Times* [30], respectively.

Geographic Analysis

Twitter Geographic Analysis

We examined the pure volume of tweets, including both engagement and original, irrespective of state population, revealing that 15.83% (15,560/98,264) of the tweets occurred in Michigan followed by New York (7765/98,264, 7.9%), District of Columbia (DC; 6822/98,264, 6.94%), California (6075/98,264, 6.18%), and New Hampshire (4262/98,264, 4.34%). However, this measure of pure tweet volume does not account for differences in population sizes across jurisdictions. **Figure 3** incorporates data from the 2010 US census to provide a normalized comparison of tweets per 100,000 people by state. Once normalized, DC rose to 1133 tweets per 100,000, nearly 4 times that of New Hampshire with 323 tweets per 100,000. The remaining three top-ranking states were Michigan, Vermont, and Rhode Island with 157, 112, and 56 tweets per 100,000, respectively.

Table 2 highlights the engagement ratio for the top 5 jurisdictions with the highest number of tweets per 100,000 population. Among the 5 jurisdictions with the highest PFAS-related Twitter volume per 100,000 population, New Hampshire displayed the highest engagement ratio (2.71) followed by Rhode Island (2.52), DC (1.23), Michigan (1.2), and Vermont (0.9).

Figure 3. Choropleth of tweets per 100,000 people by state.

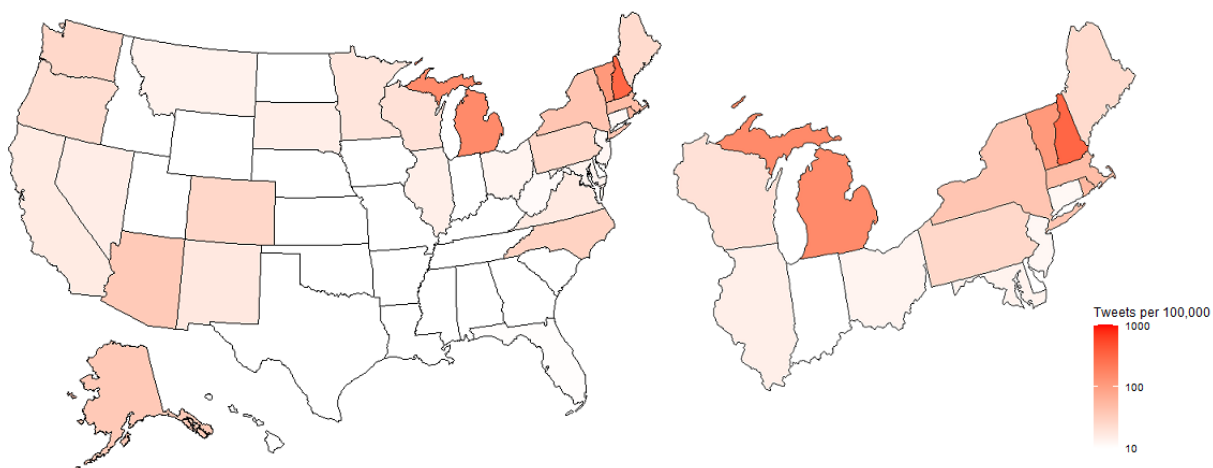


Table 2. Top 5 jurisdictions with the highest number of tweets per 100,000 population and engagement ratio.

Jurisdiction	Tweets per 100,000 population, N	Original posts, N	Engagement posts, N	Engagement ratio
District of Columbia	1133.74	3057	3765	1.23
New Hampshire	323.74	1149	3113	2.71
Michigan	157.43	7059	8501	1.2
Vermont	112.99	372	335	0.9
Rhode Island	56.15	168	423	2.52

Reddit Geographic Analysis

Of the 3126 posts in the Reddit data set, 443 (14.17%) contained jurisdictional information, allowing for geographic analysis. The remaining 85.83% (2683/3126) did not have any information for geographic linking. The top five states for PFAS-related posts by pure volume, including both original and engagement posts, were Michigan (206/443, 46.5%), North Carolina (35/443, 7.9%), Pennsylvania (28/443, 6.3%), Ohio (19/443, 4.3%), and Minnesota (14/443, 3.2%). The top jurisdictions by Reddit posts per 100,000 population were

Michigan (2.08), Vermont (1.12), Alaska (0.84), Delaware (0.56), New Hampshire (0.46), North Carolina (0.37), Minnesota (0.26), and Wisconsin (0.25; Figure 4).

Table 3 highlights the engagement ratio for the top jurisdictions with the highest number of Reddit posts per 100,000 population. Minnesota displayed the highest engagement ratio (6) followed by North Carolina (4), Michigan (3.29), Vermont (2.50), and Wisconsin (1.80). The engagement ratios for Alaska, Delaware, and New Hampshire could not be computed (N/A in Table 3) as the number of original posts was 0 for these states.

Figure 4. Choropleth of Reddit posts per 100,000 people by state.

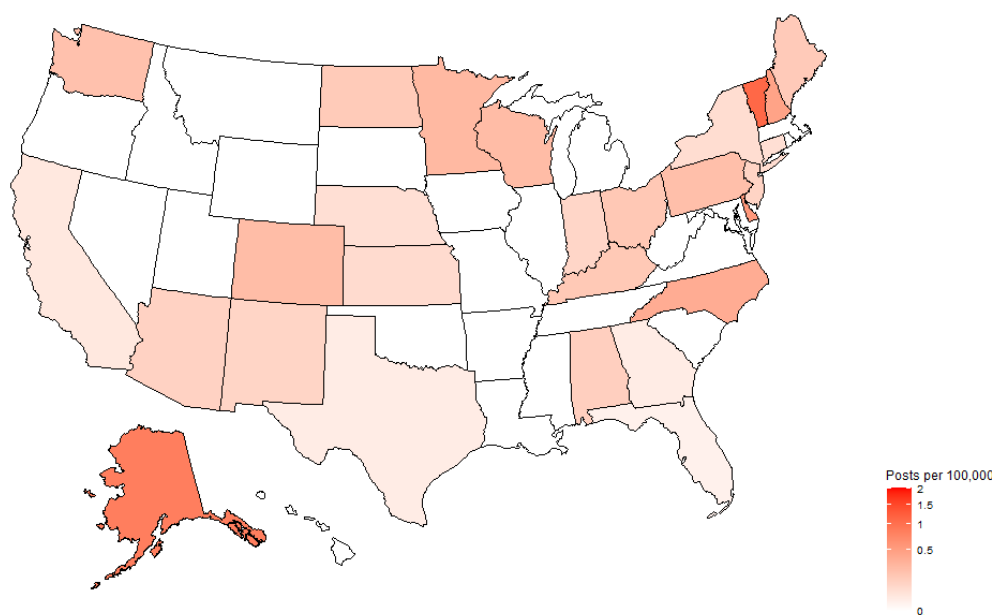


Table 3. States with the highest number of Reddit posts per 100,000 population and engagement ratio.

Jurisdiction	Reddit posts per 100,000 population	Original posts, N	Engagement posts, N	Engagement ratio
Michigan	2.08	48	158	3.29
Vermont	1.12	2	5	2.50
Alaska	0.84	0	6	N/A ^a
Delaware	0.56	0	5	N/A
New Hampshire	0.46	0	6	N/A
North Carolina	0.37	7	28	4
Minnesota	0.26	2	12	6
Wisconsin	0.25	5	9	1.80

^aN/A: not applicable.

Thematic Analysis

We identified the following three topics as the most prevalent themes, ordered from most to least commonly mentioned: (1) social media users discuss PFAS as an immediate public health concern, (2) social media users desire clarity on government agencies' roles related to PFAS, and (3) social media users discuss general mistrust of PFAS-related information.

Social Media Users Discuss PFAS as an Immediate Public Health Concern

We defined this theme, hereafter abbreviated as *PFAS are a health concern*, by the volume of posts discussing the health impacts of PFAS. Some of the thematic excerpts included a discussion of the products known to contain a PFAS derivative and their associated health effects. Examples of keywords associated with this theme included *hazard*, *crisis*, *threat*, *exposure*, *dangerous*, *unsafe*, and *risk*. Many social media users believe that PFAS are dangerous human-made chemicals that contribute to wider environmental and health issues. A significant portion of the posts associated with this theme included dialogue on the discovery of new PFAS contamination sites across the country, *breaking* news articles or studies reporting new locations with high concentrations of PFAS, and concerns for potential adverse health effects from personal exposure to PFAS. In addition, social media posts containing this theme often implicitly or explicitly expressed a lack of clarity on PFAS health effects or uncertainty on where to access relevant health information. Discussions often included questions for clarification or requested links to PFAS-related information.

Social Media Users Desire Clarity on Government Agencies' Roles Related to PFAS

We defined this theme, hereafter abbreviated as *clarity on government agencies' roles*, by the volume of posts mentioning government agencies (eg, state and local health government and federal agencies) and discussing the need for government leadership for PFAS-related community outreach and dissemination of standardized information. Examples of keywords associated with this theme included *enforce*, *state*, *regulate*, *mitigate*, and *protect*. Social media users largely engaged with PFAS discussions as it related to a desire for more government participation at the local, state, and federal levels.

Users frequently connected PFAS-related issues and concerns to contaminated water and drinking water supplies and called on the government for more active support in producing evidence-based science to identify and protect the public from further PFAS exposure. The public's confusion and uncertainty on various agencies' roles in PFAS-related activities provides an opportunity for government agencies to ensure their messaging outlines how the actions they are taking will protect public health.

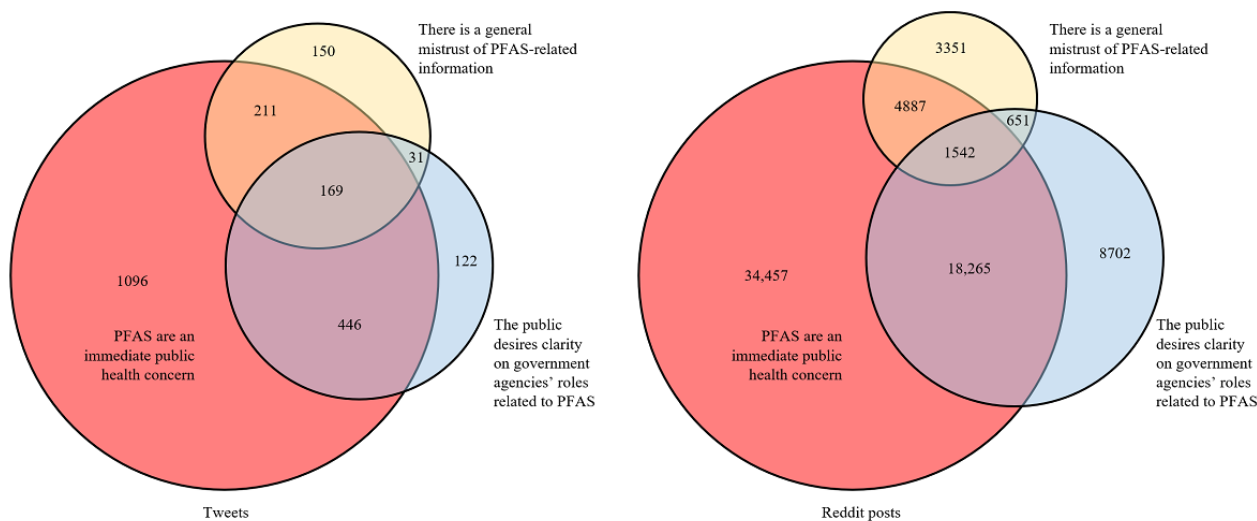
Social Media Users Discuss General Mistrust of PFAS-Related Information

We defined this theme, hereafter abbreviated as *mistrust of PFAS information*, by the volume of posts discussing mistrust of the available information related to PFAS. Examples of keywords associated with this theme included *corrupt*, *greed*, *litigation*, and *accountable*. In the last decade, there has been an exponential increase in scientific literature on the health effect of PFAS and many ongoing changes in PFAS information, in part because of jurisdictional attempts and successes in changing health exposure thresholds [3,5,10]. The data suggest that many social media users felt unable to determine the credibility of certain sources of PFAS information, particularly when conflicting information is presented by media outlets and various sources. Many users commented on the potential for political bias in journalism, often looking for authoritative sources of scientific information. Social media users often sought to expand their knowledge by discussing PFAS-related information with other users and expressed an overarching frustration based on the need to piece together accurate PFAS information.

Twitter Thematic Analysis

Of the 98,264 tweets, 59,151 (60.2%) received the *PFAS are a health concern* code, 29,160 (29.68%) received the *clarity on government agencies' roles* code, and 10,431 (10.62%) received the *mistrust of PFAS information* code. Of the 98,264 total tweets, 71,855 (73.12%) received *at least* one thematic code, and the remaining 26,409 (26.88%) received no code. As seen in Figure 5, approximately 47.33% (46,510/98,264) received exactly 1 code, and 24.22% (23,803/98,264) received 2 codes. Approximately 1.57% (1542/98,264) received all 3 codes. See the *Thematic Analysis* section above for clarification of theme definitions.

Figure 5. Overlap of tweets or Reddit posts containing at least one thematic code (Twitter: N=98,264; Reddit: N=3126). PFAS: per- and polyfluoroalkyl substance.



Reddit Thematic Analysis

Of the 3126 posts in the Reddit data set, 1922 (61.48%) received the *PFAS are a health concern* code, 768 (24.57%) received the *clarity on government agencies' roles* code, and 561 (17.95%) received the *mistrust of PFAS information* code. Of the complete Reddit data set of 3126 posts and comments, 2225 (71.18%) received at least one code, and 901 (28.82%) received no code. See the *Thematic Analysis* section above for clarification of theme definitions.

As shown in [Figure 5](#), approximately 43.76% (1368/3126) received exactly 1 code, and 22% (688/3126) received 2 codes. Approximately 5.41% (169/3126) received exactly 3 codes.

Mixed Methods Analysis

Posts that received the *PFAS are a health concern* code were 1.59 times as likely as other PFAS-related posts to be categorized as engagement (ie, a reply, comment, or retweet that responded to or engaged with an existing comment) rather than volume (ie, an original post that did not reply, comment, or retweet a previously existing comment). Posts that received the *clarity on government agencies' roles* code (odds ratio 1.35,

95% CI 1.31-1.39) and *mistrust of PFAS information* (odds ratio 1.36, 95% CI 1.3-1.42) code were also significantly more likely to be categorized as engagement rather than volume ([Table 4](#)).

Post hoc tests showed that posts related to the theme *PFAS are a health concern* were significantly more likely to be labeled as engagement posts compared with posts categorized into the other 2 themes. However, the themes *clarity on government agencies' roles* and *mistrust of PFAS information* were statistically equal in their relative levels of engagement.

The number of *likes* for a post ranged from 0 to 1447 within the data set. As shown in [Table 5](#), a post categorized as *PFAS are a health concern* was 1.78 times as likely to receive *likes* compared with other PFAS tweets. Tweets categorized as *mistrust of PFAS information* were 1.35 times as likely to receive *likes*. However, the *clarity on government agencies' roles* theme received 94% as many *likes* (incident rate ratio=0.94; $P=.001$) when compared with other PFAS-related tweets. Furthermore, when conducting 2-way post hoc comparisons across themes, we found that all themes were statistically significantly different from one another in terms of how often they received *likes* by Twitter users.

Table 4. Logistic regression results of engagement on thematic categories.

Variable	Values			
	Estimate (SE)	Z	P value	OR ^a (95% CI)
Intercept	0.12 (0.01)	11.26	<.001	— ^b
PFAS ^c are a health concern	0.46 (0.01)	34.11	<.001	1.59 (1.54-1.63)
Mistrust of PFAS information	0.31 (0.02)	13.65	<.001	1.36 (1.3-1.42)
Clarity on government agencies' roles	0.3 (0.01)	20.12	<.001	1.35 (1.31-1.39)

^aOR: odds ratio.

^bOdds ratio not provided for intercept.

^cPFAS: per- and polyfluoroalkyl substance.

Table 5. Poisson regression results of likes on thematic categories.

Variable	Values			
	Estimate (SE)	Z	P value	IRR ^a (95% CI)
Intercept	-2.25 (0.02)	-135.40	<.001	0.11 (0.1-0.11)
PFAS ^b are a health concern	0.57 (0.02)	31.09	<.001	1.78 (1.71-1.84)
Mistrust of PFAS information	0.30 (0.02)	12.74	<.001	1.35 (1.29-1.41)
Clarity on government agencies' roles	-0.06 (0.02)	-3.27	.001	0.94 (0.91-0.98)

^aIRR: incident rate ratio.

^bPFAS: per- and polyfluoroalkyl substance.

Discussion

Social Media Data Usability for Public PFAS Interest Analysis

There were several challenges identified in the collection, processing, and comparison of data associated with the original 7 disparate social media platforms we sought to evaluate. Owing to data access restrictions on Nextdoor, we could not collect any PFAS-related mentions at all on the platform nor could we capture aggregated or descriptive statistics of PFAS mentions from the platform. On Pinterest, we ran into limitations collecting PFAS mentions with required metadata (eg, time and regions across the United States), which removed the data's potential comparative value. Attempts at data collection on YouTube and Imgur did not reveal substantial user mentions of PFAS-related key terms during the period of analysis. The full sample from YouTube included <300 videos, and Imgur featured a single post.

Although Facebook can be a valuable data source for analyzing social media use, Facebook only enables data to be collected from preidentified public pages. Therefore, we determined that it was not appropriate to include Facebook data in this untargeted analysis that aimed to explore and analyze the general patterns and uses of social media for PFAS-related topics. Given these challenges, we focused our analyses on two platforms—Twitter and Reddit—based on similarities in metadata and relatively large sample sizes.

Use Trends Across Social Media Platforms

Social media data provide a lens to understand perceptions about rising public health phenomena (see the *Limitations* section for further discussion of social media user population representativeness). Comparing the availability of PFAS-related data and metadata on different social media platforms and the Twitter and Reddit findings, our analysis indicates variation in post volume by platform. Our research over the 2-year period shows increasing public dialogue on PFAS. Although there are several other platforms that can contain meaningful insights into public perceptions of PFAS, for this analysis, Twitter and Reddit were the most appropriate for understanding geographic trends and theme sentiment. Although Facebook is the most widely used social media network [35] and can be a valuable data source for analyzing social media use, Facebook only enables data to be collected from preidentified public pages. Therefore, Facebook data can be suitable for a targeted research

approach whereby researchers first select specific public Facebook pages associated with topics of interest and then collect data from those pages. This data retrieval restriction limits the use of Facebook for broad identification of related content across the entire platform and detecting general use trends.

The rising volume of posts across Twitter and Reddit and the growing visibility of PFAS across local and national news suggest an increasing interest in PFAS-related information. Considering each of the social media sources initially examined for this study, Twitter was the predominant platform for PFAS-related posts, likely owing in part to its ease of use in sharing ideas, real-time information, and trending news. However, despite having a relatively larger proportion of original posts, users engaged more frequently with Reddit posts, suggesting a more interactive platform. On the basis of our content analysis, Reddit users often provided their impressions, opinions, and perceptions of PFAS-related topics with other users. Individuals using Reddit to discuss PFAS tended to engage in follow-up discussions with the goal of gaining additional factual or anecdotal information or to provide additional resources. These differences in posts across social media platforms provide an initial understanding of the ways in which government agencies can further investigate PFAS sentiment among social media users. Through the themes that this research identified, our findings suggest that social media posts can help provide insight into some portions of the public's accurate or inaccurate understanding of PFAS-related information and determine some gaps in public knowledge.

Geographic and Event Trends

Understanding the prevalence of PFAS-related posts by jurisdiction provides a measure for PFAS public health professionals, organizations, and federal agencies to provide tailored intervention information, remediation, and community outreach by jurisdiction. The geographic data revealed a pattern of high-volume PFAS-related posts among jurisdictions affected by PFAS contamination during the time frame. Our analysis suggests that individuals living in jurisdictions with recent PFAS events appeared to be more engaged with web-based discussions that included continuous and active coverage of PFAS events by local news outlets. DC, New Hampshire, and Michigan were the top-ranked jurisdictions based on posts per 100,000 across both platforms. Much of the social media dialogue associated with Michigan and New Hampshire was related to PFAS

contamination incidents or sites such as Parchment in Michigan and Pease International Tradeport in New Hampshire.

Thematic Trends

Our analysis revealed that many posts described the contaminants as a public health threat and national crisis, especially among those living in PFAS-affected communities. As more individuals become aware of and are affected by PFAS, there is a growing sense of urgency to understand the health impacts and protective behaviors associated with PFAS exposure. This is exemplified through the predominant theme found in our analysis—*PFAS are a health concern*. Our findings suggest that both Twitter and Reddit were used to question, share, engage with, and react to evolving information related to discovered PFAS contamination. This is underscored by individual desires for access to concrete and reliable evidence-based information about PFAS and their impacts on human health and the environment.

Our results showed a notable overlap between *PFAS are a health concern* and *clarity on government agencies' roles*. Some posts stressed the notion that scientific, academic, and environmental communities could play a positive and significant role in shaping future PFAS responses and interventions by acting both as scientific experts and trustworthy resources for answering outstanding questions on both health and environmental risks. These findings present an opportunity for government agencies to provide scientific information, answer questions, and correct misperceptions. American social media users are particularly interested in understanding what is being done to mitigate exposure, the potential human health effects, and obtaining information on tangible steps individuals can take to manage their own risk. Although there is still much to be known about PFAS, our findings suggest that there is an unmet need among the public for clarity on who is leading the charge in scientific discoveries, cleanup and protection, and dissemination of scientific health-related information. Given the core theme *clarity on government agencies' roles*, government agencies have an opportunity to play a lead role in developing and distributing materials to the public on what is known about PFAS detection and to effectively communicate what remains unknown. This will not only provide PFAS information but also demonstrate the agencies' roles as it relates to PFAS and health. Furthermore, by capitalizing on social media's interactive format, government agencies can communicate information in these forums in a way that reflects audience preferences, stimulates conversation on relevant critical topics, and provides content that can be tailored to each platform's application.

On the basis of our findings, individuals and communities negatively affected by PFAS feel uninformed and subsequently excluded from proposed or in-progress solutions. There is a strong desire on social media for a trusted source of PFAS-related health information on everything from breaking news and events to longer-term response plans.

Our analysis found that PFAS content is best created with a lens that acknowledges sensitivities within current public perceptions. It is preferred that outreach efforts and scientific findings ensure clarity in word choices and accurate framing of hot topics to mitigate disinformation and encourage social media users'

engagement. Dissemination strategies may be developed through partnerships with trusted scientific and academic organizations to dispel misinformation and demonstrate collaborative active leadership on the topic. In this way, continuous and ongoing evaluation of PFAS-related social media discussions may be advantageous, especially as new discussions arise across social media platforms and further communication materials are developed and deployed nationally. On social media platforms, government agencies can demonstrate real-time awareness of current PFAS-related issues and provide diverse content to address geographically specific concerns. By better understanding social media user interests, concerns, opinions, and perceptions of PFAS, environmental health professionals will be better equipped to uncover innovative ways to inform and engage citizens on important environmental health topics.

Limitations

There are several limitations to our analysis. First, differences may exist between social media users and the general population [36], so our findings should not be extrapolated to the general population. Second, given the jurisdictional patterns found when sorting by our engagement ratio measure, we believe this measure has several inherent flaws. Most notably, it indicates instability among small population jurisdictions, as seen with several small states that were among the top 5 when measured by engagement ratio but not among the top 5 when measured by original post volume and original post volume by 100,000 people. In addition, our Reddit data were substantially truncated because of the lack of geographic indicators in the metadata. This resulted in a small sample size and more volatility in the geographic analysis across this platform. As events were determined post hoc, there was an inherent collinearity between the geography and event measures, resulting in potentially inflated SEs and very little improvement in the predictive value of the outcome. The location-masking tools and decisions to opt out of location information may also have had an impact on the geographic analysis. The length of posts allowed by Twitter also differed from what is allowed by Reddit, which limited our ability to compare results between platforms. Third, we must acknowledge the potential existence of social media *bots*—an autonomous account or network of accounts that posts and shares content according to specific predetermined rules—within the evaluated data sets that was not an immediate focus of this research effort but nevertheless may have had an impact on the spread of PFAS-related information [37]. Finally, data generated by social media platforms create unique research challenges, primarily as social media platforms create an exceptionally dynamic data environment. Specifically, across social media platforms, content is being created, disseminated, reacted to, and interacted with constantly and in real time. At any moment, a new or returning user can view a video, like a post, unlike a post, post a comment, or delete a previous comment. Thus, this dynamic communication environment presents challenges to those conducting research with its data as the data represent a finite snapshot in time rather than a continuous cumulative capture.

Conclusions

On the basis of the analysis of social media posts and activity, we find that PFAS are perceived as an immediate public health concern and there is a growing sense of urgency to understand this emerging contaminant and its potential health impacts on social media. Twitter is the predominant platform among the 7 investigated social media platforms for PFAS-related posts based on the pure volume of key terms mentioned during the 2-year time frame of the analysis. However, despite Twitter having a relatively larger proportion of original posts, Reddit posts were more frequently engaged with by users, suggesting a more interactive rather than passive platform. Some social media users seek to understand long- and short-term health risks

and access reliable PFAS information to make personal decisions that mitigate health risks. All of this underscores an opportunity for a robust public health response. Government agencies can continue using social media research to better understand the changing public sentiment on PFAS and the critical topics of interest among affected communities and then use social media as a forum for dispelling misinformation, communicating scientific findings, and providing resources for relevant public health services. Through these findings on geographic and event-based trends in PFAS-related discussions, government agencies and other partner organizations are better equipped to disseminate targeted PFAS-related scientific information and conduct community outreach.

Acknowledgments

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry. The use of trade names is for identification only and does not imply endorsement by the Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry.

Authors' Contributions

HT, CG, LL, AP, GV, and AG conceived and designed the study. AG and BG collected the data. AG, BG, and HT were involved in data analysis. HT, CG, LL, AP, GV, AG, BG, EH, CMR, and RDR were involved in interpreting the results. EH, AG, and BG drafted the initial manuscript. All authors provided critical comments and edited, revised, and approved the manuscript in its final form for submission.

Conflicts of Interest

None declared.

References

1. Blum A, Balan SA, Scheringer M, Trier X, Goldenman G, Cousins IT, et al. The madrid statement on poly- and perfluoroalkyl substances (PFASs). *Environ Health Perspect* 2015 May;123(5):107-111 [FREE Full text] [doi: [10.1289/ehp.1509934](https://doi.org/10.1289/ehp.1509934)] [Medline: [25932614](https://pubmed.ncbi.nlm.nih.gov/25932614/)]
2. Scheringer M, Trier X, Cousins IT, de Voogt P, Fletcher T, Wang Z, et al. Helsingør statement on poly- and perfluorinated alkyl substances (PFASs). *Chemosphere* 2014 Nov;114:337-339 [FREE Full text] [doi: [10.1016/j.chemosphere.2014.05.044](https://doi.org/10.1016/j.chemosphere.2014.05.044)] [Medline: [24938172](https://pubmed.ncbi.nlm.nih.gov/24938172/)]
3. Calafat AM, Wong L, Kuklennyik Z, Reidy JA, Needham LL. Polyfluoroalkyl chemicals in the U.S. population: data from the National Health and Nutrition Examination Survey (NHANES) 2003-2004 and comparisons with NHANES 1999-2000. *Environ Health Perspect* 2007 Nov;115(11):1596-1602 [FREE Full text] [doi: [10.1289/ehp.10598](https://doi.org/10.1289/ehp.10598)] [Medline: [18007991](https://pubmed.ncbi.nlm.nih.gov/18007991/)]
4. Li Y, Xu Y, Scott K, Lindh C, Jakobsson K, Fletcher T. Half-lives of PFOA, PFPeS, PFHxS, PFHpS and PFOS after end of exposure to contaminated drinking water. *Environ Epidemiol* 2019;3:237. [doi: [10.1097/01.EE9.0000608476.06577.16](https://doi.org/10.1097/01.EE9.0000608476.06577.16)]
5. Gardner P, Ellison G. Michigan's next water crisis is PFAS - and you may already be affected? MLive. 2018. URL: https://www.mlive.com/news/page/michigans_water_crisis_pfes.html [accessed 2022-02-27]
6. Basic information on PFAS. Environmental Protection Agency. 2018. URL: <https://www.epa.ie/our-services/monitoring--assessment/waste/chemicals/pfas/> [accessed 2022-02-25]
7. Perfluoroalkyl and Polyfluoroalkyl Substances (PFAS): frequently asked questions. Agency for Toxic Substances of Disease Registry. 2018. URL: https://www.atsdr.cdc.gov/pfas/docs/pfas_fact_sheet.pdf [accessed 2022-02-25]
8. Tucker DK, Macon MB, Strynar MJ, Dagnino S, Andersen E, Fenton SE. The mammary gland is a sensitive pubertal target in CD-1 and C57Bl/6 mice following perinatal perfluorooctanoic acid (PFOA) exposure. *Reprod Toxicol* 2015 Jul;54:26-36 [FREE Full text] [doi: [10.1016/j.reprotox.2014.12.002](https://doi.org/10.1016/j.reprotox.2014.12.002)] [Medline: [25499722](https://pubmed.ncbi.nlm.nih.gov/25499722/)]
9. NTP monograph on immunotoxicity associated with exposure to perfluorooctanoic acid (PFOA) or perfluorooctane sulfonate (PFOS). National Toxicology Program (NTP), U.S. Department of Health and Human Services. 2016. URL: https://ntp.niehs.nih.gov/ntp/ohat/pfoa_pfos/pfoa_pfosmonograph_508.pdf [accessed 2022-02-25]
10. Maximum contaminant level recommendation for perfluorooctanoic acid in drinking water. New Jersey Drinking Water Quality Institute. 2017. URL: <https://www.nj.gov/dep/watersupply/pdf/pfoa-recommend.pdf> [accessed 2018-10-26]

11. Grandjean P. Delayed discovery, dissemination, and decisions on intervention in environmental health: a case study on immunotoxicity of perfluorinated alkylate substances. *Environ Health* 2018 Jul 31;17(1):62 [FREE Full text] [doi: [10.1186/s12940-018-0405-y](https://doi.org/10.1186/s12940-018-0405-y)] [Medline: [30060739](https://pubmed.ncbi.nlm.nih.gov/30060739/)]
12. Cordner A, De La Rosa VY, Schaidler LA, Rudel RA, Richter L, Brown P. Correction: guideline levels for PFOA and PFOS in drinking water: the role of scientific uncertainty, risk assessment decisions, and social factors. *J Expo Sci Environ Epidemiol* 2019 Oct 29;29(6):861 [FREE Full text] [doi: [10.1038/s41370-019-0134-5](https://doi.org/10.1038/s41370-019-0134-5)] [Medline: [30926894](https://pubmed.ncbi.nlm.nih.gov/30926894/)]
13. The devil we know. IMDb. 2018. URL: <https://www.imdb.com/title/tt7689910/> [accessed 2022-02-27]
14. Boyd D, Ellison N. Social network sites: definition, history, and scholarship. *J Comput-Mediat Commun* 2008;13:210-230 [FREE Full text] [doi: [10.1111/J.1083-6101.2007.00393.x](https://doi.org/10.1111/J.1083-6101.2007.00393.x)]
15. Blank G, Reisdorf BC. The participatory web: a user perspective on Web 2.0. *Inform Commun Society* 2012 May;15(4):537-554. [doi: [10.1080/1369118x.2012.665935](https://doi.org/10.1080/1369118x.2012.665935)]
16. Fung IC, Tse ZT, Fu K. The use of social media in public health surveillance. *Western Pac Surveill Response J* 2015;6(2):3-6 [FREE Full text] [doi: [10.5365/WPSAR.2015.6.1.019](https://doi.org/10.5365/WPSAR.2015.6.1.019)] [Medline: [26306208](https://pubmed.ncbi.nlm.nih.gov/26306208/)]
17. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. *PLoS One* 2010 Nov;5(11):e14118 [FREE Full text] [doi: [10.1371/journal.pone.0014118](https://doi.org/10.1371/journal.pone.0014118)] [Medline: [21124761](https://pubmed.ncbi.nlm.nih.gov/21124761/)]
18. Da'ar OB, Yunus F, Md Hossain N, Househ M. Impact of Twitter intensity, time, and location on message lapse of bluebird's pursuit of fleas in Madagascar. *J Infect Public Health* 2017 Jul;10(4):396-402 [FREE Full text] [doi: [10.1016/j.jiph.2016.06.011](https://doi.org/10.1016/j.jiph.2016.06.011)] [Medline: [27423931](https://pubmed.ncbi.nlm.nih.gov/27423931/)]
19. Thackeray R, Neiger BL, Smith AK, Van WS. Adoption and use of social media among public health departments. *BMC Public Health* 2012;12:242 [FREE Full text] [doi: [10.1186/1471-2458-12-242](https://doi.org/10.1186/1471-2458-12-242)] [Medline: [22449137](https://pubmed.ncbi.nlm.nih.gov/22449137/)]
20. McCausland K, Maycock B, Leaver T, Jancey J. The messages presented in electronic cigarette-related social media promotions and discussion: scoping review. *J Med Internet Res* 2019 Feb 05;21(2):e11953 [FREE Full text] [doi: [10.2196/11953](https://doi.org/10.2196/11953)] [Medline: [30720440](https://pubmed.ncbi.nlm.nih.gov/30720440/)]
21. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying topics for e-cigarette user-generated contents: a case study from multiple social media platforms. *J Med Internet Res* 2017 Jan 20;19(1):e24 [FREE Full text] [doi: [10.2196/jmir.5780](https://doi.org/10.2196/jmir.5780)] [Medline: [28108428](https://pubmed.ncbi.nlm.nih.gov/28108428/)]
22. 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](https://doi.org/10.2196/publichealth.8641)] [Medline: [29263018](https://pubmed.ncbi.nlm.nih.gov/29263018/)]
23. Huo J, Desai R, Hong Y, Turner K, Mainous AG, Bian J. Use of social media in health communication: findings from the health information National Trends Survey 2013, 2014, and 2017. *Cancer Control* 2019 Apr 18;26(1):1073274819841442 [FREE Full text] [doi: [10.1177/1073274819841442](https://doi.org/10.1177/1073274819841442)] [Medline: [30995864](https://pubmed.ncbi.nlm.nih.gov/30995864/)]
24. Djuricich AM. Social media, evidence-based tweeting, and JCEHP. *J Contin Educ Health Prof* 2014;34(4):202-204. [doi: [10.1002/chp.21250](https://doi.org/10.1002/chp.21250)] [Medline: [25530289](https://pubmed.ncbi.nlm.nih.gov/25530289/)]
25. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2019. URL: <https://www.R-project.org/> [accessed 2022-02-27]
26. Strauss A, Corbin J. Grounded theory methodology: an overview. In: Denzin NK, Lincoln YS, editors. *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage Publications; 1994:273-284.
27. Bernard HR, Wutich A, Ryan GW. *Analyzing Qualitative Data: Systematic Approaches*. Thousand Oaks, CA: SAGE Publications; 2016.
28. Lustgarten A. How the EPA and the Pentagon Downplayed a Growing Toxic Threat. 2018. URL: <https://www.propublica.org/article/how-the-epa-and-the-pentagon-downplayed-toxic-pfas-chemicals> [accessed 2022-02-27]
29. Bailey D. Democrats allege corruption in Snyder administration for PFAS water contamination. WZZM. 2018. URL: <https://tinyurl.com/yc5vf4bn> [accessed 2022-02-27]
30. Turkewitz J. Toxic 'Forever chemicals' in drinking water leave military families reeling. *The New York Times*. 2019. URL: <https://www.nytimes.com/2019/02/22/us/military-water-toxic-chemicals.html> [accessed 2022-02-27]
31. Lipton E, Turkewitz J. Pentagon pushes for weaker standards on chemicals contaminating drinking water. *The New York Times*. 2019. URL: <https://www.nytimes.com/2019/03/14/us/politics/chemical-standards-water-epa-pentagon.html> [accessed 2022-02-27]
32. Herrick RL, Buckholz J. Polyfluoroalkyl substance exposure in the Mid-Ohio River Valley, 1991–2012. *Environ Pollut* 2017;28:50-60 [FREE Full text] [doi: [10.1016/j.envpol.2017.04.092](https://doi.org/10.1016/j.envpol.2017.04.092)]
33. Boronow KE, Brody JG, Schaidler LA, Peaslee GF, Havas L, Cohn BA. Serum concentrations of PFASs and exposure-related behaviors in African American and non-Hispanic white women. *J Expo Sci Environ Epidemiol* 2019 Mar;29(2):206-217 [FREE Full text] [doi: [10.1038/s41370-018-0109-y](https://doi.org/10.1038/s41370-018-0109-y)] [Medline: [30622332](https://pubmed.ncbi.nlm.nih.gov/30622332/)]
34. Holden E. Trump administration condemned over delaying action on toxic drinking water. *The Guardian*. 2019. URL: <https://www.theguardian.com/environment/2019/feb/14/epa-toxic-chemicals-pfos-pfoa-delay> [accessed 2022-02-27]
35. Most popular social networks worldwide as of July 2020, ranked by number of active users. Statista. 2020. URL: <https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/> [accessed 2022-02-27]

36. Blank G. The digital divide among Twitter users and its implications for social research. *Soc Sci Compu Rev* 2016 Oct 12;35(6):679-697. [doi: [10.1177/0894439316671698](https://doi.org/10.1177/0894439316671698)]
37. Broniatowski D, 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](https://doi.org/10.2105/ajph.2018.304567)]

Abbreviations

DC: District of Columbia
EPA: Environmental Protection Agency
PFAS: per- and polyfluoroalkyl substance
PFOA: perfluorooctanoic acid
PFOS: perfluorooctane sulfonic acid

Edited by JMIRPE Office; submitted 09.11.20; peer-reviewed by AUR Bacha, J Raubenheimer; comments to author 09.12.20; revised version received 01.02.21; accepted 27.12.21; published 11.03.22.

Please cite as:

Tian H, Gaines C, Launi L, Pomales A, Vazquez G, Goharian A, Goodnight B, Haney E, Reh CM, Rogers RD
Understanding Public Perceptions of Per- and Polyfluoroalkyl Substances: Infodemiology Study of Social Media
J Med Internet Res 2022;24(3):e25614
URL: <https://www.jmir.org/2022/3/e25614>
doi: [10.2196/25614](https://doi.org/10.2196/25614)
PMID: [35275066](https://pubmed.ncbi.nlm.nih.gov/35275066/)

©Hao Tian, Christy Gaines, Lori Launi, Ana Pomales, Germaine Vazquez, Amanda Goharian, Bradley Goodnight, Erica Haney, Christopher M Reh, Rachel D Rogers. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 11.03.2022. 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.

Original Paper

The Public Perception of the #GeneEditedBabies Event Across Multiple Social Media Platforms: Observational Study

Congning Ni¹, ME; Zhiyu Wan^{1,2,3}, PhD; Chao Yan^{1,2}, PhD; Yongtai Liu¹, MS; Ellen Wright Clayton^{3,4,5,6}, JD, MD; Bradley Malin^{1,2,3,7}, PhD; Zhijun Yin^{1,2,3}, PhD

¹Department of Computer Science, Vanderbilt University, Nashville, TN, United States

²Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

³Center for Genetic Privacy & Identity in Community Settings, Vanderbilt University Medical Center, Nashville, TN, United States

⁴Center for Biomedical Ethics and Society, Vanderbilt University Medical Center, Nashville, TN, United States

⁵Department of Pediatrics, Vanderbilt University Medical Center, Nashville, TN, United States

⁶School of Law, Vanderbilt University, Nashville, TN, United States

⁷Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, United States

Corresponding Author:

Zhijun Yin, PhD

Department of Biomedical Informatics

Vanderbilt University Medical Center

2525 West End Avenue

Nashville, TN, 37232

United States

Phone: 1 6159363690

Email: zhijun.yin@vumc.org

Abstract

Background: In November 2018, a Chinese researcher reported that his team had applied clustered regularly interspaced palindromic repeats or associated protein 9 to delete the gene *C-C chemokine receptor type 5* from embryos and claimed that the 2 newborns would have lifetime immunity from HIV infection, an event referred to as #GeneEditedBabies on social media platforms. Although this event stirred a worldwide debate on ethical and legal issues regarding clinical trials with embryonic gene sequences, the focus has mainly been on academics and professionals. However, how the public, especially stratified by geographic region and culture, reacted to these issues is not yet well-understood.

Objective: The aim of this study is to examine web-based posts about the #GeneEditedBabies event and characterize and compare the public's stance across social media platforms with different user bases.

Methods: We used a set of relevant keywords to search for web-based posts in 4 worldwide or regional mainstream social media platforms: Sina Weibo (China), Twitter, Reddit, and YouTube. We applied structural topic modeling to analyze the main discussed topics and their temporal trends. On the basis of the topics we found, we designed an annotation codebook to label 2000 randomly sampled posts from each platform on whether a supporting, opposing, or neutral stance toward this event was expressed and what the major considerations of those posts were if a stance was described. The annotated data were used to compare stances and the language used across the 4 web-based platforms.

Results: We collected >220,000 posts published by approximately 130,000 users regarding the #GeneEditedBabies event. Our results indicated that users discussed a wide range of topics, some of which had clear temporal trends. Our results further showed that although almost all experts opposed this event, many web-based posts supported this event. In particular, Twitter exhibited the largest number of posts in opposition (701/816, 85.9%), followed by Sina Weibo (968/1140, 84.91%), Reddit (550/898, 61.2%), and YouTube (567/1078, 52.6%). The primary opposing reason was rooted in ethical concerns, whereas the primary supporting reason was based on the expectation that such technology could prevent the occurrence of diseases in the future. Posts from these 4 platforms had different language uses and patterns when they expressed stances on the #GeneEditedBabies event.

Conclusions: This research provides evidence that posts on web-based platforms can offer insights into the public's stance on gene editing techniques. However, these stances vary across web-based platforms and often differ from those raised by academics and policy makers.

KEYWORDS

CRISPR/Cas9; gene-edited babies; social media; stance learning; text mining; content analysis

Introduction

Background

The clustered regularly interspaced palindromic repeats (CRISPR) or CRISPR-associated protein 9 (Cas9) genetic scissor is a revolutionary gene-editing technology for the life sciences, for which its creators were awarded the 2020 Nobel Prize in Chemistry [1]. However, much remains unknown about the long-term effects of applying CRISPR in *human gene editing* (HGE), and experts in ethics and policy have particularly cautioned against its use in *germline gene editing* (GGE) to influence the genetic code of a human embryo [2,3]. However, in November 2018, a Chinese researcher, Jiankui He, reported that his team applied this tool to delete the *gene C-C chemokine receptor type 5* from embryos and claimed that the 2 newborns would benefit from lifelong immunity from HIV infection. This event, which is often referenced as #GeneEditedBabies in the web-based domain, has raised widespread concerns in the scientific community about HGE, especially GGE, for numerous reasons, which include but are not limited to biological safety and ethical implications [4-6].

The experiments of Jiankui He violated Chinese regulations and ignored international ethical norms [7]. Most criticisms against his actions concentrated on assertions that the experiments (1) were unnecessary as existing treatment strategies can sufficiently block HIV paternal transmission without the risk of GGE [8] and (2) violated the protocols required for general genetic procedures [9], expressing a concern that off-target editing can lead to dangerous and unpredictable mutations [10,11]. To expedite global cooperation on science and its governance, an international commission comprising experts from 10 countries released a guidance report in September 2020 on how to determine whether such an application is sufficiently well-developed for clinical use [12].

However, the worldwide debate has focused primarily on the perspective of academics and policy makers. By contrast, little is known about how the public, who are often influenced by geographic region and local culture, perceived this event, as well as these issues more generally. Gaining intuition into the public's perspectives on these matters is critical to understanding social norms and expectations for potential beneficiaries and victims. In this respect, the public's perspective is a core component in determining resource allocation, political policies, and participation in clinical trials, all of which influence the development and adoption of such technologies as they evolve [13,14].

To date, research on public attitudes has mainly applied survey methods to investigate hypothetical personal opinions regarding gene-editing techniques [15-18], and few have considered this specific actual event. At the same time, web-based social media platforms have become an essential medium for billions of people to learn about and comment on trending events, including

#GeneEditedBabies, in a timely manner, making them an ideal resource to study public perception. Web-based social media platforms accommodate speech from wide demographics, and the discussions and expressions therein are not controlled by predefined questionnaires. Thus, they may effectively reduce the bias in the data [19]. As a result, over the past decade, user-generated content from web-based social media platforms has been increasingly relied upon to study the public's attitudes regarding a broad range of topics, including weight stigmatization [20], antivaccination [21], lung cancer screening [22], the Brexit referendum [23], and political debates [24]. A total of 3 recent studies of web-based data investigated the #GeneEditedBabies event; however, they were limited in that they either focused only on sentiment [25] or relied on data from only 1 web-based platform [26,27], which, as we show in this study, paints an incomplete and likely biased picture.

Objective

In this work, we present a substantially broader investigation into the public's perceptions of the #GeneEditedBabies event through the lens of four popular social media platforms known to be influenced by different cultures and demographics: *Sina Weibo* (based in China), *Twitter*, *Reddit*, and *YouTube*. Specifically, we apply topic modeling to >220,000 posts to analyze what had been discussed about this event across these platforms and examine how these topics changed over time. Our results indicate that the main topics were related to news; the technique itself; posting words on the web; and discussions from the perspectives of disease, risk, laws, ethics, and scientific literature. We also observed that certain topics have clear temporal trends based on the heat of the relative discussions. On the basis of the topic analysis results, we further design a codebook to annotate 2000 randomly selected posts from each platform, which we use to investigate how web-based posts supported, opposed or held a neutral stance on this event. The results indicate that the public's web-based posts had much more divided stances toward this event than toward those of academics, although the former varied across platforms with different language patterns.

The findings of this study indicate that society can learn about the public perception of controversial events using web-based data; however, we must be careful about drawing conclusions from any single platform.

Methods

Overview

Our research focused on the data that were collected from four popular and publicly accessible web-based platforms with user bases from different regions and backgrounds: *Sina Weibo*, *Twitter*, *YouTube*, and *Reddit*. The internal review board at our university designated this project as non-human subjects research and exempt from review. All posts presented in this

paper have been rephrased for privacy concerns and demonstration purposes.

Data Preparation

As the #GeneEditedBabies event was instigated by a Chinese scientist, it is natural to examine how people reacted to this event on Sina Weibo, the most popular microblogging platform with 500 million users in China [28]. The data from this platform have been applied to investigate a broad range of topics, including suicide prediction [29] and the mining of characteristics of patients with COVID-19 in China [30].

Twitter and YouTube are web-based platforms with a broader international basis, where approximately 20% of their users are from the United States, and the other 80% are from other countries [31]. Owing to its instant posting feature, Twitter has been recognized as an ideal platform for tracking people's comments or attitudes toward social events over time [32,33]. YouTube, a web-based video-sharing platform that enables users to upload, view, rate, and comment on videos has also become a valuable resource for studying the public's stances [34,35].

Reddit is a social news aggregation, web content rating, and discussion website, with approximately 50% of users from the United States [36]. Reddit users are more likely to write longer posts and generate deeper discussions on specific topics than users from other platforms. In addition, most Reddit users reportedly have a college degree [37].

We acknowledge that the users who mentioned the #GeneEditedBabies event do not necessarily represent the general population in each platform and, thus, may not represent the same demographics as those previously mentioned. However, because of the wide coverage of their user base, we believe that the posts on these platforms can provide intuition into the public's stance regarding the #GeneEditedBabies event in a cross-regional and cross-cultural manner [38].

To collect data, we defined a set of keywords that were related to the #GeneEditedBabies event: *jiankui he*, *gene-edited babies*, *crisper babies*, 贺建奎, 基因编辑婴儿, and their variations. Given that each platform has distinct application programming interface (API) use policies and that we initiated data collection 3 days after the event (November 29, 2018), we relied upon several different strategies to collect data. Specifically, we used the Twitter streaming API to continuously collect data from the Twitter stream and used the Twitter search API to fetch the data that were generated before the initial data collection date. For the other web-based platforms, we used their search APIs to collect data. We collected data from November 26, 2018, to October 19, 2019; aggregated the data on each platform; and removed duplicates. For Twitter, YouTube, and Reddit, we focused only on posts written in English.

Topic Analysis

To characterize what has been discussed in the #GeneEditedBabies event and provide a basis for the following stance annotation, we conducted a topic analysis of all collected data using structural topic modeling (STM) [39]. STM is a topic modeling framework that can leverage document-level

meta-information (eg, publication date and author) to improve inference and qualitative interpretability. However, because of language differences, it is challenging to generate topics by applying a single model to all the data. In addition, as many posts and comments in YouTube data do not have a well-defined posting date (eg, *one month ago*), we could not accurately build temporal trends for this platform's topics. As a result, we applied STM on Twitter and Reddit data sets together to (1) extract general topics, (2) analyze topic trends, and (3) compare topic differences; we applied STM on Sina Weibo and YouTube data sets to extract only general topics separately. We believe that temporal analysis can help understand how related news triggered the public's discussions on web-based social media platforms.

We removed stop words, punctuation, and numbers and replaced words with their stems before applying STM. As STM is based on unsupervised learning, we relied on exclusivity and semantic coherence to select the optimal number of topics K^* by running STM, with each candidate for the number of topics ranging from 5 to 25. Exclusivity corresponds to the distinctiveness of words with the highest probabilities in the topics, whereas semantic coherence measures the probability of co-occurrence of these words in a topic. As large exclusivity often results in small semantic coherence and vice versa, an ideal choice of K^* should strike a balance between these 2 measures. To do so, we first selected the 3 topic number candidates for K^* that exhibited the largest proportion of topics, with scores within 1 SD of the mean for the metrics of interest. The mean score was calculated for all the candidate topic numbers. We then manually compared and selected the number of topics that resulted in the most interpretable topics through the authors' observations and discussions.

Stance Annotation

Overview

Topic analysis can provide a high-level picture of what has been discussed in general. However, it cannot precisely tell whether people support or oppose this event. Therefore, after topic analysis, we conducted a stance analysis to further investigate how web-based posts from different platforms reacted to this event.

Coding Question

To identify and analyze the stance of web-based users in the #GeneEditedBabies event, we randomly selected 2000 posts from each platform for annotation. We designed a codebook for data annotation based on reading, independently summarizing, and collectively discussing the topic analysis results among the authors (CN, ZW, CY, and ZY) of this manuscript. We formulated 2 annotation tasks for each selected post. First (question 1 [Q1]), we asked what the post's stance was toward this event and let annotators select one of the following options: *no clear stance*, *support*, *oppose*, or *neutral*. As the distinction between *no clear stance* and *neutral* stances can be somewhat nuanced, we instructed annotators that a post with a *neutral* stance must explicitly balance between supporting and opposing stances.

If the answer to Q1 was different from *no clear stance*, we then asked annotators to perform a second task (question 2 [Q2]) to identify what issues the post raised when expressing its stance. Q2 was a nonexclusionary multi-option question, in which annotators were instructed to choose any of the following options based on the instructions of the codebook: (1) *techniques* (eg, considering the development of science or technologies),

(2) *ethics*, (3) *laws*, (4) *judging Jiankui He* (eg, on his education, appearance, or ambition), (5) *judging the organization* (eg, He’s institute or country), (6) *others* (eg, those not listed above, but the missing aspect should be recorded), and (7) *no reasons are mentioned*. Table 1 presents several examples along with their answers.

Table 1. Examples of posts with answers to the questions posed in the annotation tasks.

Post examples	Platform	Q1 ^a choice (suggested)	Q2 ^b choice (suggested)
“年轻有为，反正我不会这个技术。(He is young and promising; I do not understand this technology anyway)”	Sina Weibo	Support	Techniques and judging Jiankui He
“The first gene-edited babies claimed in China—The Mainichi https://t.co/uNL0QFfdur ”	Twitter	No clear stance	N/A ^c
“I mean Nazi scientists went to hell in a tote basket but produced some of the most influential research of the 20th century.”	Reddit	Neutral	Techniques and ethics
“It sounds like any other mad scientist story.”	YouTube	Oppose	Judging Jiankui He

^aQ1: question 1.

^bQ2: question 2.

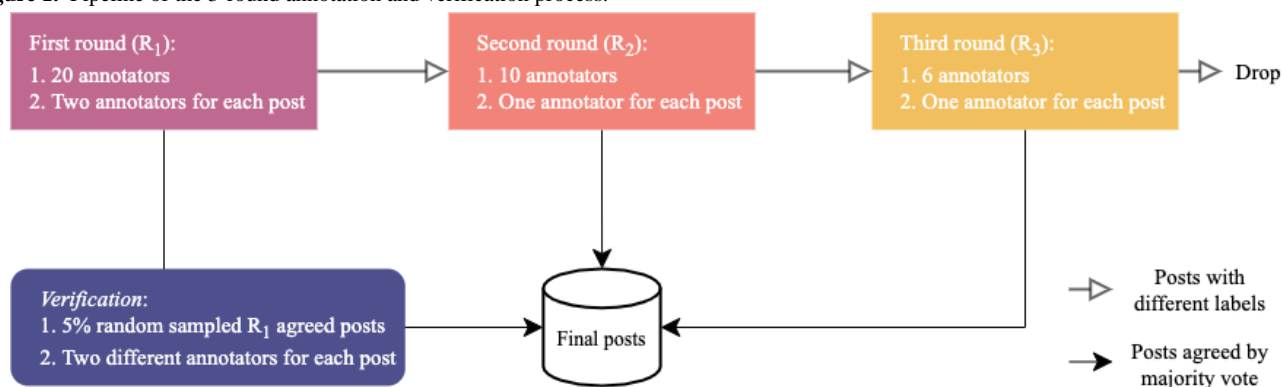
^cN/A: not applicable.

Annotation Strategies

As the languages used in the 4 platforms are either Chinese or English, we recruited 20 college students who were fluent in both languages to annotate the data. We selected bilingual annotators to reduce bias in stance judgment induced by differences in the language spoken. Before assigning the task,

we provided the codebook and tutorial examples to train the annotators and discuss with them. Once they were comfortable with the requirements and the task, we proceeded to formal data annotation. As stance is subjective, to ensure the annotation quality, we conducted 3 rounds of *verification* for Q1, the single option question. Figure 1 shows the annotation and verification processes.

Figure 1. Pipeline of the 3-round annotation and verification process.



A total of 8000 posts from 4 platforms, with 2000 posts per platform, were labeled in our annotation. In the first round (R_1), we randomly partitioned 20 annotators into 10 groups of 2 annotators each and assigned each group to 800 selected posts (200 from each platform). After annotation, we compared the labeling results within the groups. When 2 annotators labeled a post with the same stance, the task for the post was considered complete. We calculated the inner-annotator agreement between the 2 annotators in the first round using the Cohen κ score.

Instead of requiring 2 annotators to select the same options in Q2, which would be very challenging, we calculated a weighted score for each option based on their answers as follows:

$$s_{ij} = \frac{1}{K} \sum_{k \in A_k} \dots$$

s_{ij} represents the score of option o_j in post p_i , i is the post number from 1 to 800, j is the option number from 1 to 7 (see the abovementioned coding questions), A_k represents the set of options that annotator k chooses, and K is the total number of annotators who labeled post p_i :

$$K = \dots$$

We sent all posts with conflicting stance labels into a second round of annotation (R_2) for tie-breaking purposes. In this round, we selected another 10 annotators as adjudicators. If a post received 3 different stance labels in 2 rounds of annotation, we sent it to the third round of annotation (R_3) for further tie breaking. Multiple rounds of annotation can reduce manual effort and improve efficiency. However, some posts received

2 stance labels in R_1 , whereas some other posts received 4 stance labels in R_3 . It is possible that the posts that received the same stance labels in R_1 could obtain different labels from 2 (different) annotators, which may have made the annotations unreliable. To examine the extent to which such cases existed, we randomly selected 5% of such posts and assigned them to 2 annotators who had yet to label them for a verification process.

For posts with stances settled beyond R_1 , we extracted the answers to Q2 from the 2 annotators who determined their final stance label. Once we calculated the weighted score $\frac{1}{M} \sum_{i=1}^M s_i$ for each option o_j in each post p_i , from Q2, we estimated $\frac{1}{M} \sum_{i=1}^M p_i$, the probability that an issue (corresponding to option o_j) was raised in a web-based platform regarding a particular stance. $\frac{1}{M} \sum_{i=1}^M p_i$ was calculated as follows:

$$\frac{1}{M} \sum_{i=1}^M p_i$$

Where M represents the total number of labeled posts for a platform, and L represents the number of options in Q2.

Stance Language Analysis

After comparing the stance differences across the 4 platforms, we continued to examine how the language used in the posts differentiated supporting and opposing stances across the platforms. To perform this analysis, we focused on annotated data with only supporting or opposing stances. We removed stop words and extracted the lemma form of each word. We then calculated the saliency [40] $s(w)$ for a word w in each platform as follows:

$$s(w) = \frac{p(w|t)}{p(w)}$$

In the abovementioned equation, t refers to either a supporting or opposing stance, and p represents the probability of w within the data corpus. The larger the saliency value, the more informative the word is for differentiating among stances. We retained the 50 most informative words for each platform. For each selected word w , we used the relevance score [41] to determine how informative it was to a stance t , which is defined as follows:

$$r(w, t) = \frac{\lambda \cdot f(w, t)}{f(w, t) + (1 - \lambda) \cdot \max_{t'} f(w, t)}$$

In the abovementioned equation, λ is used to control the trade-off between the word frequencies within a stance and the word importance in distinguishing 2 stances. Here, a small λ highlights rare but exclusive words for a stance, whereas a large λ highlights frequent but not necessarily exclusive words for the stance. In this study, we set $\lambda=0.6$, as suggested by Sievert and Shirley [41].

Results

Data Preparation

Table 2 summarizes the data collected for this study. Within the first round of annotation, the inner-annotator agreement was 0.44 (SD 0.11), which indicated moderate agreement. The median number of posts published by each author was 1 for all platforms, suggesting that most users only had one post or comment in our data set.

Table 2. A summary of the data set collected for this study^a.

Platform and users	Post type	Posts	Posts per user, mean (median) ^b	Post length, mean (SD) ^c
Sina Weibo				
2800	Microblog	4941	1.8 (1)	81.4 (76.8)
83,265	Comment	131,126	1.6 (1)	16.6 (17.5)
Twitter				
24,960	Tweet	47,147	1.9 (1)	14.1 (4.2)
Reddit				
866	Submission	3205	3.7 (1)	71.0 (324.6)
11,678	Comment	22,417	1.9 (1)	43.3 (65.9)
YouTube				
31,237	Comment	48,172	1.5 (1)	36.1 (60.4)

^aOn the basis of the properties of each platform, the posting type varies.

^bRepresents the average number (median) of posts per user in each platform.

^cRepresents the average word count (SD) of each post.

Topic Analysis

Topics on Twitter and Reddit

On the basis of our criterion, we set K^* to 15 for the topic analysis. We refer the reader to Multimedia Appendix 1, Tables S1-S3, for further details. Table 3 reports on the 15 extracted

topics; the most representative words; expected topic proportions (the probability of each topic in all the posts); and the 5 summarized themes, which were summarized by authors (CN and ZY) by manually reviewing the posts with the highest probability, and the difference in topical prevalence between

Twitter and Reddit. The most representative words per topic were based on the rank ordering of their probabilities.

Table 3. Topics, top words, and the 5 summarized themes.

Theme and topic	Top representation words	ETP ^a	Difference (SD) ^b
Communication			
11	<i>just, realli, happen, get, your, like, one, actual, yeah, and guy</i>	0.066	-0.092 (0.001)
9	<i>think, that, thing, dont, right, doesnt, cant, anyth, bad, and good</i>	0.059	-0.090 (0.001)
5	<i>can, even, genet, point, know, well, someth, understand, way, and still</i>	0.058	-0.078 (0.001)
6	<i>daddi, [curse word #1], look, cock, [curse word #2], littl, long, talk, perfect, and want</i>	0.038	-0.039 (0.001)
Discussions			
15	<i>use, diseas, crispr, cell, mutat, technolog, cancer, ccr, techniqu, and cure</i>	0.071	0.013 (0.001)
1	<i>gene, edit, hiv, human, embryo, risk, may, twin, genom, and studi</i>	0.064	0.010 (0.001)
14	<i>peopl, kid, child, parent, children, dont, rich, already, isnt, and wouldnt</i>	0.059	-0.099 (0.001)
13	<i>ethic, now, scientif, experi, said, done, govern, public, need, and communiti</i>	0.049	-0.035 (0.001)
7	<i>will, year, futur, chang, come, human, today, -izumi, time, and permalinksavecontextful</i>	0.047	-0.019 (0.001)
8	<i>alphian, omegian, planet, civil, ampxb, galaxi, low-level, omega, cosmic, and betian</i>	0.017	-0.016 (<0.001)
Gene editing			
12	<i>geneedit, genomeedit, crisprca, crispr, tool, geneediting..., genetherapi, amp, genet, and genom</i>	0.121	0.184 (0.002)
News			
3	<i>gene-edit, scientist, chines, babi, claim, crispr, report, jiankui, scandal, and miss</i>	0.144	0.168 (0.002)
4	<i>babi, news, gene-edit, first, moratorium, creat, world, controversi, world, and nobelist</i>	0.089	0.074 (0.002)
10	<i>china, crisprbabi, research, scienc, confirm, jiankui, halt, stanford, investig, and say</i>	0.080	0.066 (0.001)
Web-based posting			
2	<i>comment, pleas, post, question, automat, thank, remov, bot, moder, and thought</i>	0.038	-0.046 (0.001)

^aETP: expected topic proportion.

^bDifference represents the difference in the prevalence of the topic between Twitter and Reddit, where a positive (negative) value suggests a topic was more frequently discussed on Twitter (Reddit). All the differences were significant with $P < .001$ according to a 2-tailed paired t test.

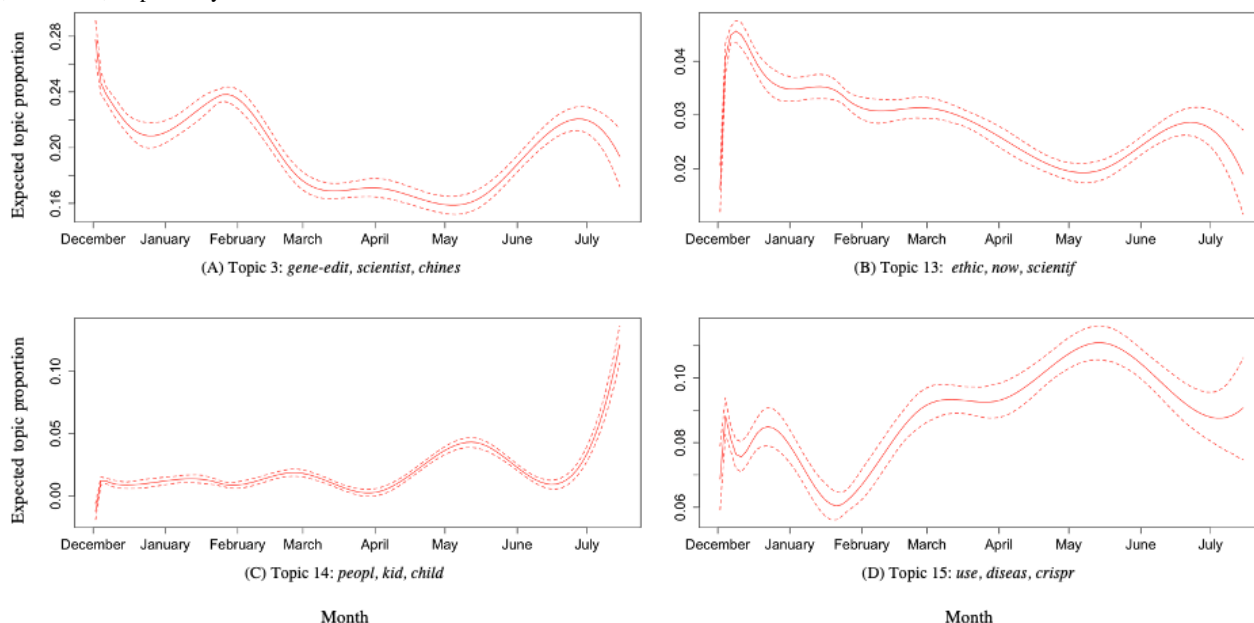
Five themes were generated by the authors: *communications* (topics refer to regular web-based chatting), *discussions* (topics related to the discussion of this event), *gene editing* (topics of gene editing technique), *news* (topics of #GeneEditedBabies), and *posting on the web* (topic of common web-based communities' words). Topics with positive (negative) scores were more often discussed on Twitter (Reddit). For example, it can be seen that topic 12 from the *gene-editing* theme exhibited the highest score, followed by topics 3, 4, and 10, which are all part of the *news* theme. This suggests that Twitter is more dedicated to the sharing of news. By contrast, topic 14 exhibited the lowest negative score, a topic in the *discussion* theme about people and children. Topics 11, 9, and 5 of the

communication themes also exhibited negative scores, which suggests that they were more likely to be mentioned on Reddit as well. The above findings clearly demonstrate the different topics on the 2 platforms.

Topic Temporal Trend Analysis

Figure 2 illustrates the temporal trends for the 4 representative topics derived from Twitter and Reddit. We interpreted the trends by manually examining the posts with the largest probabilities for each topic during a specific period. Note that the scale of the y-axis for each topic that we selected in Figure 2 varies because of different expected topic proportions (Table 3).

Figure 2. Topic temporal trends for 4 topics generated by applying structural topic modeling on Twitter and Reddit data. The solid line represents the mean expected topic proportion and the dashed lines represent 1 SD. The x-axis corresponds to the posting time, whereas the y-axis corresponds to the expected topic proportion obtained from structural topic modeling. (A) to (D) represent the monthly changes in the expected topic proportion of Topic 3, 13, 14 and 15, respectively.



Topic 3 (in Figure 2) mainly refers to the scientist Jiankui He in this event. After the beginning of the event, the popularity of this topic decreased sharply but then experienced an increase in February 2019. This increase corresponded in time to a rumor that *Chinese gene-editing scientist missing amid rumors of arrest: report says Chinese scientist He Jiankui has been arrested* on Twitter. He again attracted attention in the news in June 2019 because of an announcement by Denis Rebrikov at the Kulakov National Medical Research Center in Moscow, who was planning to repeat his experiment once he obtained official approval [42].

Topic 13 communicates the public's concerns about the event (eg, legal issues). This topic experienced a decreasing trend as the #GeneEditedBabies event became outdated. However, it experienced a peak around May 2019 because of an article published in the journal *Nature: CRISPR babies: when would the world be ready?* [43], which stirred a discussion on Reddit.

The temporal trend for topic 14, which focuses on human rights and was more likely to be discussed on Reddit, exhibited an increasing trend, with a local peak during May 2020 when a Reddit user shared an article with the title of *What we risk as humans if we allow gene-edited babies: a philosopher's view* [44], which received >600 comments. The popularity of this topic continued to rise rapidly in July 2020, perhaps because of the news of *Five couples lined up for CRISPR babies to avoid deafness* [45], which received numerous retweets on Twitter.

Topic 15, which focuses on gene-editing technology, grew over time, with a peak around May 2019. The peak of its popularity appeared to coincide with news related to how gene editing helps cure disease. For example, 1 such news story posted on Twitter in May 2019 was *Researcher Brian Liau and his team combine CRISPR gene-editing technology with chemical profiling to tease out acute myeloid leukemia mechanisms* [18].

Topics on YouTube and Sina Weibo

We also applied STM to YouTube and Sina Weibo data sets to generate $K^*=15$ topics using the same criteria for processing Twitter and Reddit data sets, respectively. The three most popular topics on YouTube were summarized as visual arts (*video*, *gattaca*, and *episod*), thoughts (*want*, *get*, and *think*), and human evolution (*human*, *speci*, and *evolut*). The top three popular topics on Sina Weibo were news (*事情 event*, *科研 research*, and *国际 international*), science (*科学 science*, *科学家 scientist*, and *成果 result*), and concerns (*法律 law*, *胚胎 embryos*, and *人体 human*). Other topics on YouTube included mentions of race and culture (keywords: *tan*, *skin*, *cultur*, *hair*, and *racist*), and general thoughts (*peopl*, *talk*, *feel*, and *understand*). By contrast, Sina Weibo included topics presenting strong negative emotions, such as strong opposition to the event (*出生 born*, *死 death*, *可怜 pity*, *毁灭 destroy*, *人性 humanity*, and *危害 danger*) and the discussion on inequity between the rich and poor (*穷人 poor*, *富人 rich*, *普通人 normal people*, and *小白鼠 experimental mice* [used as a metaphor]).

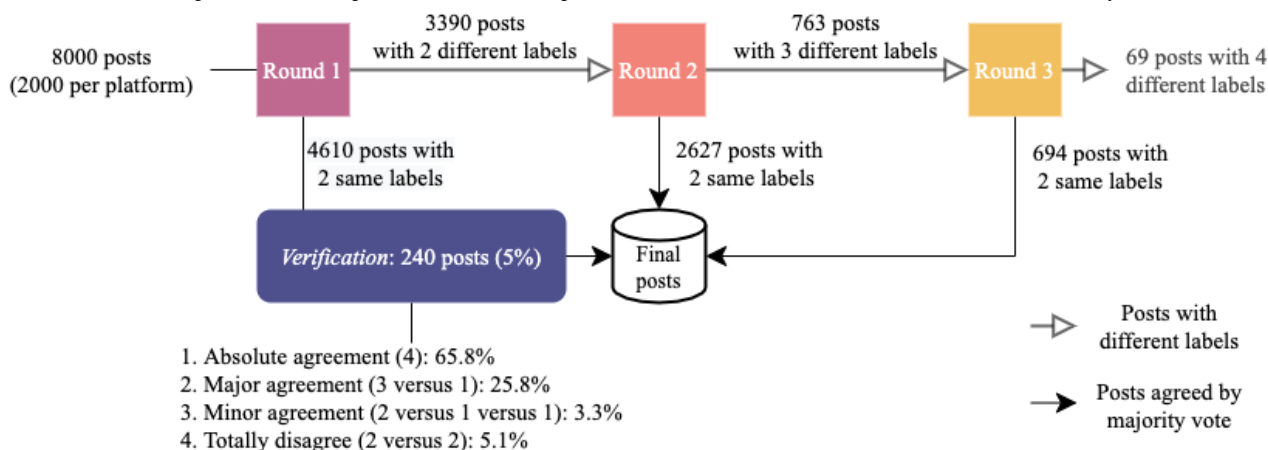
Stance Analysis

Annotation Results

Figure 3 illustrates the labeling process. Both annotators agreed in 57.62% (4610/8000) of the posts that were labeled in the first round. After the second round of annotation, 3 annotators remained in disagreement on 22.51% (763/3390). The 9% (69/763) of posts that were annotated with 4 different options were removed because of the inability to reach any consensus. Regarding verification, 65.8% (158/240) of the examined posts received the same labels as the first round (ie, absolute agreement); 25.8% (62/240) received 1 different label (ie, major agreement); 3.3% (8/240) received 2 different labels (ie, minor agreement); and only 5.1% (12/240) received 2 of the same

labels, which differed from the original, suggesting a reliable labeled data set for further analysis.

Figure 3. The annotation process for 8000 posts. Note that the 69 posts with 4 different labels were removed from further analysis.



Stance Distribution

We calculated the number and percentage of each stance in each platform among the posts that clearly had a stance in the annotated data. The percentage of removed posts with *no clear stance* was 42.8% (853/1993) on Sina Weibo, 59% (1174/1990) on Twitter, 54.4% (1070/1967) on Reddit, and 45.6% (902/1978) on YouTube. Although all 4 platforms generally contained more opposing stances than other stances, the ratios of those stances were quite different. Specifically, Sina Weibo and Twitter exhibited similar rates of opposition (approximately 84.91% [968/1140]); however, Sina Weibo’s rate of support was almost double that of Twitter (37/816, 4.5%). In contrast, Reddit and YouTube had much lower (higher) rates of opposition (support) at 61.25% (550/898; 257/898, 28.62%) and 52.6% (567/1078; 405/1078, 37.6%), respectively. Sina Weibo exhibited a lower rate of neutral stance (70/1140, 6.1%) than the other 3 platforms (all of which were around 10%). It is notable that, although both Twitter and YouTube have broader international user coverage, YouTube exhibited a more divided stance toward #GeneEditedBabies. We suspect this may be because of the fact that YouTube presents its content in videos, such that users can comment under them and share conflicting ideas in an irrelevant manner. By contrast, Twitter is more akin to a news (re)distribution center, where it is difficult to support a deep discussion.

Although a post-level analysis of stance may be biased by superusers who publish substantially more posts than others, we found that within the annotated posts with a clearly expressed stance, >95% of users had only 1 post (Sina Weibo: 1005/1057, 95.08%; Twitter: 726/755, 96.16%; Reddit: 822/858, 95.8%; and YouTube: 965/1012, 95.35%), and no users had >5 posts. This indicates that the influence of superusers on our findings based on randomly sampled data is limited.

Concern Analysis

Supporting Stance

Figure 4 shows that *techniques* were the primary consideration when posts supported the #GeneEditedBabies event on each web-based platform. The rates of Sina Weibo and YouTube posts supporting this event were slightly >50%, whereas those

for Twitter and Reddit were closer to 40%. A qualitative examination of the related posts suggests that users believed the event could help advance scientific knowledge for clinical trials using gene-editing technologies and, thus, create an opportunity to cure severe diseases and benefit humans in the future. A Sina Weibo user commented the following:

解剖学最早也是反人类的，遭到大多数人的反对，[...]，恭喜先生走出新的世界纪录。(Anatomy was also anti-human at the earliest and was opposed by most people, [...], congratulations to Mr. for stepping out of the new world record.) [post 1, Sina Weibo]

It should be noted that *no reason* was the second most selected option for Q2 on YouTube (76.5/467, 16.4%), Sina Weibo (23/114.5, 20.1%), and Twitter (14.5/38.5, 37.7%) and the third most selected option for Reddit (51/311, 16.4%). The difference in *no reason* ranking within the platforms may be as Reddit discussions are more likely to compose longer posts than the discussions on the other 3 web-based platforms (Table 1) and, thus, have more of a chance to share reasons in their posts. Ethics was another critical consideration when posts supported this event on Reddit (66.5/310.5, 21.4%), YouTube (73/467, 15.6%), and Sina Weibo (17/114.5, 14.8%). The users in these posts did not believe that this event lacked ethical concerns. Instead, these users intimated that the event could help patients with genetic diseases. The two posts indicated the following:

It’s not morality, it’s cowardice to seek permission from 30 governing bodies before taking every step. [...] No risk, no reward. [Post 2, Reddit]

Nice work! [...] If you can help parents with a genetic disorder have a healthy baby, I don’t understand why people are so upset. [Post 3, YouTube]

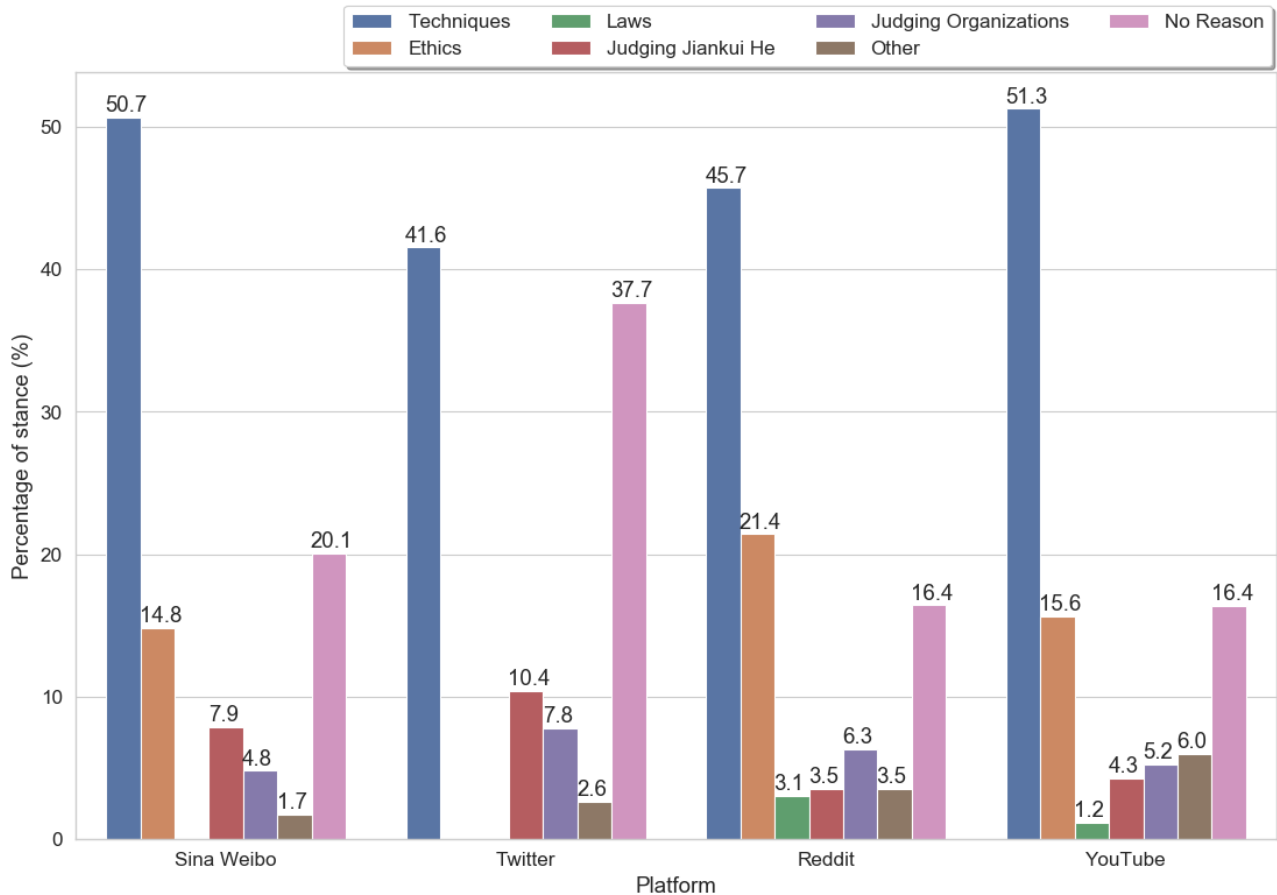
We also found posts expressing a supporting stance directly based on the aspects of Jiankui He or organizations (eg, He’s institute or country); posts on Sina Weibo and Twitter tended to talk about Jiankui He himself, whereas Reddit and YouTube tended to say more of his organizations. The following two posts offer examples in these categories:

总有人要付出代价。如果你不迈出第一步，后面的人就永远不敢前进。(There is always someone who has to pay the price. If you don't take the first step, the ones behind will never dare to move forward.) [Post 4, Sina Weibo]

China will leave us behind because we refuse to edit human genes like this. [Post 5, YouTube]

Some Reddit and YouTube posts supported this event as they believed it could help push forward new strategies and regulations for gene therapy, as indicated in post 3.

Figure 4. The reasons for supporting stances within 4 platforms.



Opposing Stance

Figure 5 illustrates the distribution of the issues identified in each platform for the opposing stance. Although there were many comments without any reason, especially on Twitter, ethics was the primary concern for Reddit (259/724, 35.8%), YouTube (235/714, 32.9%), and Sina Weibo (376.5/1258.5, 29.92%). A YouTube opposer said the following:

Who owns the rights to products and technologies? [...] I don't think it's been adequately addressed, nor is there transparency in the scientific community about it. [Post 6, YouTube]

The fact that the technique is relatively new and could lead to many unknown and serious issues was another consideration in Reddit (209/724, 28.9%), YouTube (204/714, 28.6%), and Sina Weibo (179/1258.5, 14.22%). A Redditor wrote the following:

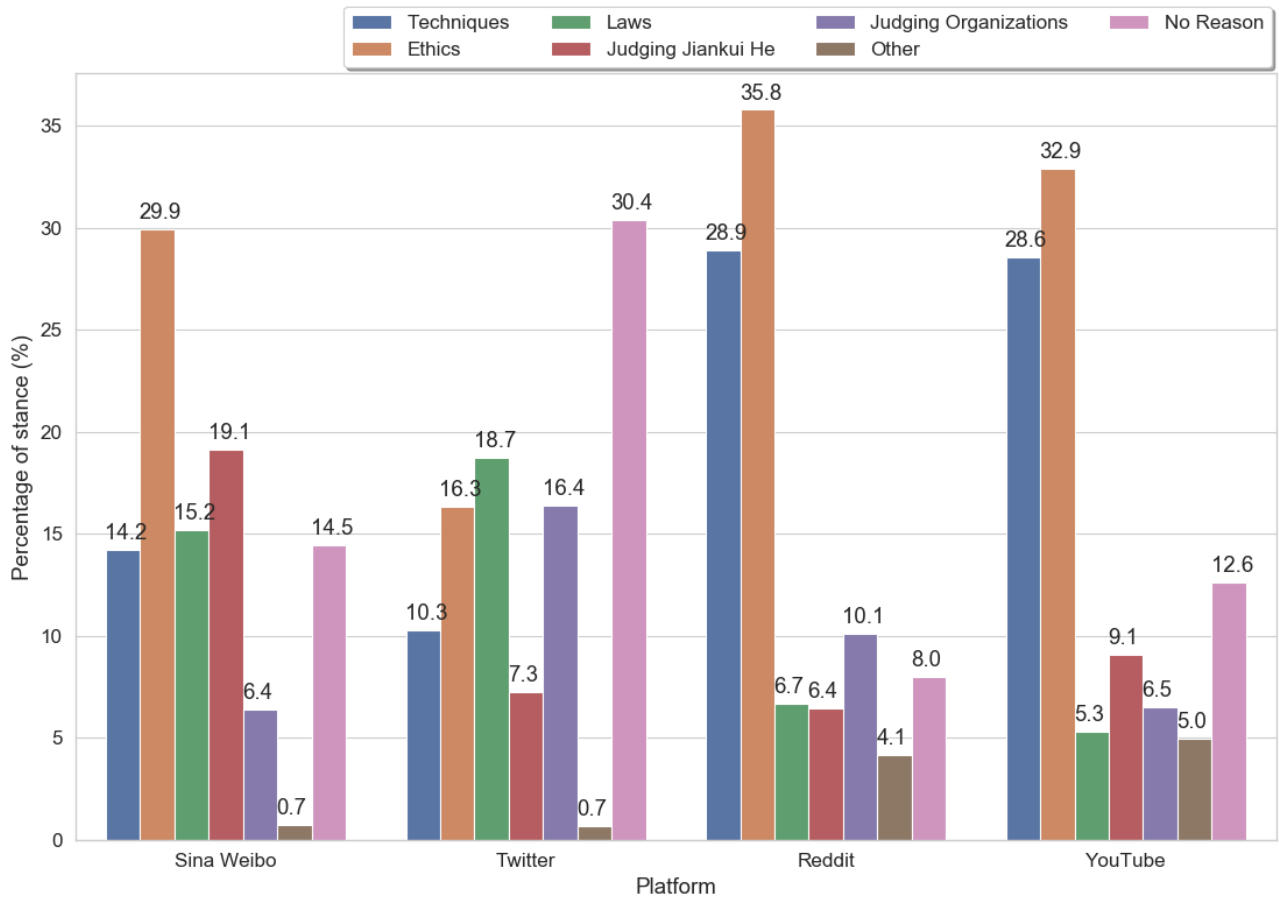
Genes are complex. So, if it doesn't work, it's as if you're not fighting the infection properly and there's a higher risk of dying. [post 7, Reddit]

However, Sina Weibo had more posts and insults (240.5/1258.5, 19.1%) that were not directed at the event or the technology but were only directed at the individual (Jiankui He). One such quote is as follows:

说白了， 还是利益熏心!! 一身的铜臭味!! (To be straightforward, (He) was too greedy and only cared about the money!) [post 8, Sina Weibo]

Notably, Twitter is quite different from the other platforms in that the major considerations mentioned in this platform were laws (151/806.5, 18.7%), organizations (132/806.5, 16.4%), and ethics (131.5/806.5, 16.3%). Techniques (83/806.5, 10.3%) had a smaller percentage than those exhibited in Sina Weibo and much less than on YouTube and Reddit. A possible explanation for this phenomenon is that many users directly tweeted the news without attaching their own commentary, whereas the discussions on the other platforms were more likely to comment under either video or other posts.

Figure 5. The reasons for opposing stances within 4 platforms.



Language Analysis

Figure 6 illustrates the word clouds for each platform, with a larger font size indicating a more informative word.

Figure 6. Salient words for supporting (green) and opposing (red) stances within each web-based platform. A larger font size suggests a higher relevance of the associated word in the corresponding stance and platform. Note that all words presented here are in their lemma form.



Sina Weibo

The most informative terms used by Sina Weibo supporters were 试管婴儿 (in vitro fertilization), 自然选择 (natural selection), 癌症 (cancer), 时代 (time), good, 日心说 (heliocentrism), 宣扬 (advocate), 骄傲 (pride), 核弹 (nuclear bomb), and 医药 (drug). These people argued that the event might be part of natural selection, and from an ethical

perspective, there might be no difference between in vitro fertilization and gene-edited babies because of human intervention. Some people believed this was good and that society should be proud as they have the potential to cure diseases such as cancer. Others argued that gene editing was similar to dual use technologies [46], such as nuclear technology, which could be used to create bombs but also be beneficial for energy development; eventually, it would be

accepted just as heliocentrism. By contrast, opposers were more likely to mention words related to law or stakeholders, such as 基因编辑 (gene edit), 广东省 (Guangdong province), 央视 (CCTV), 非法 (illegal), 明令禁止 (clearly forbidden), and 审查 (governance). It should be noted that 吐 (throw-up) is a common Chinese term in Sina Weibo that expresses an emotion in strong opposition.

Twitter

Supporters used many positive words such as *exciting*, *OK*, *innovation*, *protect*, and *enhance*. We found that male contraceptives were also mentioned as they, together with gene editing babies and other findings, were reported as the most exciting innovations of 2018, and the link to the web-based version of this report received many retweets. By contrast, people with an opposing stance used words that were more relevant to this event specifically, such as *baby*, *geneedit*, *scientist*, *china*, *illegal*, *death*, *break*, *halt*, *stop*, *rule*, and *risk*. In this regard, Twitter shared similarities with Sina Weibo in that opposers from both platforms quote news or web-based links (eg, *http*, *cnn*, *cnet* on Twitter and *http* and *CCTV* in Sina Weibo).

Reddit

Posts on Reddit were more multidimensional and tended to be more thoughtful than posts on other platforms. Supporters discussed this event from the perspective of religion, using terms such as *conservative*, and argued that they did not see any religion-related issues. They also mentioned *hereditary*. They believed that this technology could be used to cure autism or certain predispositions. They also used positive terms such as *effective* and *efficient*. In addition, these supporters likened this technique to the increase in deaths when automobiles were first introduced. The words in the supporting stance also demonstrate that Reddit posts were likely to write deep or scientific, dialectical explanations to express supporting opinions. By contrast, the opposing views included words such as *read*, *bring*, *datum*, *unknown*, and *add*, expressing their concerns about the unknown effects that such a technique might bring to the human gene pool. These opposers also used more technical terms such as *germline*, *off target*, *protein*, *allele*, *reproduction*, *deletion*, and *intervention*. Many of them believed it was dangerous, comparable with the strong opposing emotion word 吐 (*throw up*) used on Sina Weibo.

YouTube

The posts on YouTube invoked a very different language style compared with posts on other platforms. For example, supporters used the term *successful* in expressing their supporting stance and argued that gene-editing techniques will make people healthier and live longer, just as vaccines do today. Although the supporters also used terms that appear to be in opposition, such as *despite*, *greedy*, *harmful*, and *struggle*, a closer examination of the related posts showed that these posts presented 2 stances but with a solid preference for a supporting stance. By contrast, opposers on YouTube were more likely to use dramatic words such as *zombie*, *machine*, *extinction*, and *inferior* and conspiracy-related words such as *assume*, *fake*, and

lie. When talking about the law issue, these posts tended to use the word *jail*.

In summary, a comparison of informative terms shows a clear distinction in the language used in these 4 platforms. Moreover, the difference in language occurred in both the supporting and opposing stances toward this event. A possible explanation for this distinction is that it may be an artifact of user populations with different cultures or demographic backgrounds. We believe that this further supports the need to consider the public's stance from multiple diverse web-based platforms to form a more comprehensive picture. There are certainly other possible explanations, such as the fact that these differences may be because of the political stances or related screening policies of each individual platform. However, further investigation is necessary to determine whether this conjecture is correct.

Discussion

Principal Findings

This study introduced a novel, cross-cultural, and cross-regional investigation of the public's stance toward the #GeneEditedBabies event by using data from 4 popular web-based social media platforms. We found that the platforms focused on different topics and that some had clear temporal trends around the news associated with this event and related techniques. In addition, we found that opposition to this event was high in general; however, the stances varied across platforms. Among the discussions that expressed a clear stance, Twitter exhibited the largest percentage of posts (701/816, 85.9%) in opposition, followed by Sina Weibo (968/1140, 84.91%), Reddit (550/898, 61.2%), and YouTube (567/1078, 52.59%). The main reasons people supported this event were as the techniques were thought to be scientific advances that have the potential to cure debilitating genetic diseases. By contrast, the main reasons for opposition were ethical and legal concerns. The variation we observed may stem from the differences in the cultural and demographic backgrounds of the user base in each platform.

Comparison With Prior Work

Earlier Stance Analysis on Gene Editing

Our findings regarding Sina Weibo and Twitter are aligned with a conclusion in a recent survey of responses to a hypothetical, norm-violating application of gene-editing technology that such applications would *invite public backlash that can spill across national boundaries* [47]. However, as we demonstrated, Reddit and YouTube still had a substantial proportion of posts supporting this event. Other survey-based studies have also found that people tend to have supporting but cautious attitudes toward HGE. For example, 1 study showed that adults from Australia were comfortable with editing human and animal embryos but only for research purposes and to enhance human health [48]. Similarly, Japanese users generally accepted the use of genome editing for disease-related genes; however, many were concerned about its risks [49]. Although many Dutch adults responding to a survey considered the risks of GGE to be substantially greater than its benefits, they may approve of using GGE if it is sufficiently safe and effective and used for a

disease instead of enhancement [50]. In addition, another survey based on the responses from 1537 participants across 67 countries (87% of whom were White and mainly from the United States, Australia, Canada, and the United Kingdom) found that respondents generally supported GGE for medical applications, and resistance was mainly reported by people with religious beliefs or working experience in genetics [51].

However, as a recent review of nationally representative public opinion surveys summarized [52], many of these surveys did not capture the views after the #GeneEditedBabies event, *which could have raised awareness of HGE or affected views of edits among public*. In addition, the authors of the review advocated collecting more data from around the world to capture the views of different segments of populations. In this regard, our multiplatform investigation provides more insight into how people with different cultures and from different regions discussed this event, findings that can help fill the gap in this field.

Analysis of the Event Using Web-Based Data

Several studies have investigated public discussions using data from social platforms. For example, Calabrese et al [53] examined the perceptions of CRISPR on the web through the application of a semantic network analysis of Twitter messages, and Müller et al [26] analyzed tweets about CRISPR or Cas9 technology and trained machine learning models to classify tweets. However, these studies used data from Twitter only, which was not easily accessible in mainland China. Moreover, they focused on the sentiment of tweets (eg, positive or negative) and the type of subject (eg, applying the technology to humans as opposed to other organisms) instead of the stance (eg, in support or in opposition) and greater understanding of concerns (eg, why hold this stance?). Furthermore, the #GeneEditedBabies event was not the main focus. Zhang et al [27] compared the differences in language used to discuss gene editing before and after this event on Sina Weibo. Liu and Lapata [54] conducted a sentiment analysis of news reports and tweets about this event that were collected from Google, Baidu, Sina Weibo, and Twitter about this event. However, all of these investigations are fundamentally different from ours in that they selected various platforms for data collection and focused solely on sentiment analysis, which is insufficient to characterize the stance expressed in a comment.

Web-Based Multiplatform Data Analysis

Multiplatform studies are useful for reducing the bias caused by a single platform or a small amount of data. In particular, multiplatform data resources can enrich analysis and optimize the training results of classifiers [55,56]. For example, Schifanella et al [57] leveraged the contextual information carried by visuals to decode the sarcastic tone of multimodal posts by using images and texts from Instagram, Tumblr, and Twitter. In addition, most multiplatform data analyses targeted a particular topic, such as travel (eg, discussion about hospitality and tourism [58]), business (eg, the preferences and discussions of e-cigarette users [59]), and health (eg, a study of symptoms of depression and anxiety of American young adults [60]). Our study contributes to this area of research by using multiplatform data to conduct a reasonable and comprehensive analysis of

stances across various cultures, regions, and network environments.

Limitations and Future Work

However, there are certain limitations in our study that can serve as the basis for future research. First, although we focused on 4 popular publicly accessible platforms, there are other major web-based platforms (eg, Facebook), the data of which could be used to enrich the analysis. Second, the set of keywords we relied upon was limited to the names of the event and the scientist. Although we believe that these are likely to cover most of the perspectives that can be detected, we did not conduct a pretest to confirm if this was the case. Third, we calculated the percentage of stances based on randomly selected posts and showed that superusers had a limited impact on our findings. Future work may consider directly sampling all users to avoid this issue. Fourth, although our annotated data set is sufficiently large to obtain insight into the public's stance, it would be interesting to conduct stance temporal trend analysis with more labeled data. Finally, our findings reflect only the opinions of people who use web-based social media platforms to express their stances on this event. It is necessary to investigate how their perspectives relate to people who do not post on social media.

This project laid the foundation for a wide array of future studies. First, to analyze the language used, we applied a saliency word cloud to show the differences in language among posts from different stances and platforms. Alternative methods, such as semantic networks [61], could be considered in future work to investigate a different perspective of topic modeling; for example, the relationships among the words. Second, it would be useful to determine whether the public's stance on gene editing changed as a result of this event. A way by which this could be accomplished would be to compare posts about gene-editing technology before and after the #GeneEditedBabies event across social media platforms. Third, although we focused only on post-level content for analysis, it will be useful to attach conversations, such as the comments under the posts in each platform, to enrich this research.

Implications and Conclusions

This study has several notable implications. First, this study shows that web-based social media platforms can serve as efficient tools for tracking people's reactions to a series of news regarding a public event in a timely manner. However, such reactions can vary depending on the platform from which the posts come. Using multiplatform data to study public events can help to obtain a more comprehensive understanding of people's stances worldwide. Second, web-based posts exhibit more divided stances on, understandings of, and interpretations regarding a cutting edge but controversial technique than those expressed by academic professionals. This adds further weight to the need to listen to public voices and increase public engagement in policy formation beyond the scientific community [6]. Finally, the public's observed stances and the factors that web-based posts considered can both help guide the development of promotional materials to improve awareness and understanding of this technique for posts on different

platforms [62] as well as take public concerns more fully into account.

Although we focused on studying the public's stance regarding this specific event, our annotation methods and analysis

strategies can be readily adapted to investigate other social events (eg, Black Lives Matter or presidential elections) and public health promotions (eg, cancer screening or COVID-19 vaccination) in a cross-cultural or cross-regional manner.

Acknowledgments

This research was supported by the National Institutes of Health grant RM1HG009034.

Conflicts of Interest

None declared.

Multimedia Appendix 1

The top words of experiments of evaluating proper K for structural topic modeling and task examples of questions annotators got before formally starting annotation.

[[DOCX File, 41 KB - jmir_v24i3e31687_app1.docx](#)]

References

1. Advanced information. The Nobel Prize. URL: <https://www.nobelprize.org/prizes/chemistry/2020/advanced-information/> [accessed 2022-02-20]
2. Camporesi S, Marks L. Gene editing: survey invites opinions. *Nature* 2015 Nov 26;527(7579):446. [doi: [10.1038/527446c](https://doi.org/10.1038/527446c)] [Medline: [26607533](https://pubmed.ncbi.nlm.nih.gov/26607533/)]
3. Rocha LF, Braga LA, Mota FB. Gene editing for treatment and prevention of human diseases: a global survey of gene editing-related researchers. *Hum Gene Ther* 2020 Aug;31(15-16):852-862. [doi: [10.1089/hum.2020.136](https://doi.org/10.1089/hum.2020.136)] [Medline: [32718240](https://pubmed.ncbi.nlm.nih.gov/32718240/)]
4. Wang C, Zhai X, Zhang X, Li L, Wang J, Liu D. Gene-edited babies: Chinese academy of medical sciences' response and action. *Lancet* 2019 Jan;393(10166):25-26. [doi: [10.1016/s0140-6736\(18\)33080-0](https://doi.org/10.1016/s0140-6736(18)33080-0)]
5. Allyse M, Bombard Y, Isasi R, Michie M, Musunuru K, Ormond KE. What do we do now?: responding to claims of germline gene editing in humans. *Genet Med* 2019 Oct;21(10):2181-2183 [FREE Full text] [doi: [10.1038/s41436-019-0492-3](https://doi.org/10.1038/s41436-019-0492-3)] [Medline: [30914827](https://pubmed.ncbi.nlm.nih.gov/30914827/)]
6. Second international summit on human genome editing: continuing the global discussion. The National Academies Press. URL: <https://www.nap.edu/catalog/25343/second-international-summit-on-human-genome-editing-continuing-the-global-discussion> [accessed 2022-02-20]
7. Li J, Walker S, Nie J, Zhang X. Experiments that led to the first gene-edited babies: the ethical failings and the urgent need for better governance. *J Zhejiang Univ Sci B* 2019 Feb 18;20(1):32-38. [doi: [10.1631/jzus.b1800624](https://doi.org/10.1631/jzus.b1800624)]
8. Zafer M, Horvath H, Mmeje O, van der Poel S, Semprini AE, Rutherford G, et al. Effectiveness of semen washing to prevent human immunodeficiency virus (HIV) transmission and assist pregnancy in HIV-discordant couples: a systematic review and meta-analysis. *Fertil Steril* 2016 Mar;105(3):645-55.e2 [FREE Full text] [doi: [10.1016/j.fernstert.2015.11.028](https://doi.org/10.1016/j.fernstert.2015.11.028)] [Medline: [26688556](https://pubmed.ncbi.nlm.nih.gov/26688556/)]
9. Wang H, Yang H. Gene-edited babies: what went wrong and what could go wrong. *PLoS Biol* 2019 Apr;17(4):e3000224 [FREE Full text] [doi: [10.1371/journal.pbio.3000224](https://doi.org/10.1371/journal.pbio.3000224)] [Medline: [31039144](https://pubmed.ncbi.nlm.nih.gov/31039144/)]
10. Lund RJ, Närvä E, Lahesmaa R. Genetic and epigenetic stability of human pluripotent stem cells. *Nat Rev Genet* 2012 Oct;13(10):732-744. [doi: [10.1038/nrg3271](https://doi.org/10.1038/nrg3271)] [Medline: [22965355](https://pubmed.ncbi.nlm.nih.gov/22965355/)]
11. Is the CRISPR baby controversy the start of a terrifying new chapter in gene editing? *Vox*. URL: <https://www.vox.com/science-and-health/2018/11/30/18119589/crispr-gene-editing-he-jiankui> [accessed 2022-02-20]
12. Heritable Human Genome Editing. Washington, DC: The National Academies Press; 2020.
13. Drezner D. All Politics Is Global Explaining International Regulatory Regimes. Princeton, New Jersey, United States: Princeton University Press; 2008.
14. McCaughey T, Sanfilippo P, Gooden G, Budden D, Fan L, Fenwick E, et al. A global social media survey of attitudes to human genome editing. *Cell Stem Cell* 2016 May 05;18(5):569-572 [FREE Full text] [doi: [10.1016/j.stem.2016.04.011](https://doi.org/10.1016/j.stem.2016.04.011)] [Medline: [27152441](https://pubmed.ncbi.nlm.nih.gov/27152441/)]
15. Gaskell G, Bard I, Allansdottir A, da Cunha RV, Eduard P, Hampel J, et al. Public views on gene editing and its uses. *Nat Biotechnol* 2017 Nov 09;35(11):1021-1023. [doi: [10.1038/nbt.3958](https://doi.org/10.1038/nbt.3958)] [Medline: [29121022](https://pubmed.ncbi.nlm.nih.gov/29121022/)]
16. Weisberg SM, Badgio D, Chatterjee A. Corrigendum: a CRISPR new world: attitudes in the public toward innovations in human genetic modification. *Front Public Health* 2017;5:161 [FREE Full text] [doi: [10.3389/fpubh.2017.00161](https://doi.org/10.3389/fpubh.2017.00161)] [Medline: [28691005](https://pubmed.ncbi.nlm.nih.gov/28691005/)]

17. Hendriks S, Giesbertz N, Bredenoord A, Repping S. Reasons for being in favour of or against genome modification: a survey of the Dutch general public. *Hum Reprod Open* 2018;2018(3):hoy008 [FREE Full text] [doi: [10.1093/hropen/hoy008](https://doi.org/10.1093/hropen/hoy008)] [Medline: [30895249](https://pubmed.ncbi.nlm.nih.gov/30895249/)]
18. McCaughey T, Budden DM, Sanfilippo PG, Gooden GE, Fan L, Fenwick E, et al. A need for better understanding is the major determinant for public perceptions of human gene editing. *Hum Gene Ther* 2019 Jan;30(1):36-43. [doi: [10.1089/hum.2018.033](https://doi.org/10.1089/hum.2018.033)] [Medline: [29926763](https://pubmed.ncbi.nlm.nih.gov/29926763/)]
19. Holtzblatt L, Drury J, Weiss D, Damianos L, Cuomo D. Evaluating the uses and benefits of an enterprise social media platform. *J Soc Media Organ* 2013;1(1):1-21.
20. Ortiz RR, Smith A, Coyne-Beasley T. A systematic literature review to examine the potential for social media to impact HPV vaccine uptake and awareness, knowledge, and attitudes about HPV and HPV vaccination. *Hum Vaccin Immunother* 2019;15(7-8):1465-1475 [FREE Full text] [doi: [10.1080/21645515.2019.1581543](https://doi.org/10.1080/21645515.2019.1581543)] [Medline: [30779682](https://pubmed.ncbi.nlm.nih.gov/30779682/)]
21. Featherstone JD, Barnett GA, Ruiz JB, Zhuang Y, Millam BJ. Exploring childhood anti-vaccine and pro-vaccine communities on twitter – a perspective from influential users. *Online Social Networks Media* 2020 Nov;20:100105. [doi: [10.1016/j.osnem.2020.100105](https://doi.org/10.1016/j.osnem.2020.100105)]
22. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. *BMC Cancer* 2013 Oct 29;13:508 [FREE Full text] [doi: [10.1186/1471-2407-13-508](https://doi.org/10.1186/1471-2407-13-508)] [Medline: [24168075](https://pubmed.ncbi.nlm.nih.gov/24168075/)]
23. Grčar M, Cherepnalkoski D, Mozetič I, Kralj Novak P. Stance and influence of Twitter users regarding the Brexit referendum. *Comput Soc Netw* 2017;4(1):6 [FREE Full text] [doi: [10.1186/s40649-017-0042-6](https://doi.org/10.1186/s40649-017-0042-6)] [Medline: [29266132](https://pubmed.ncbi.nlm.nih.gov/29266132/)]
24. Lai M, Tambuscio M, Patti V, Ruffo G, Rosso P. Stance polarity in political debates: a diachronic perspective of network homophily and conversations on Twitter. *Data Knowl Eng* 2019 Nov;124:101738. [doi: [10.1016/j.datak.2019.101738](https://doi.org/10.1016/j.datak.2019.101738)]
25. Xinyu L, Ting W, Jieming M. Comparative sentiment analysis of Chinese and English multi-source web information of gene editing baby event. *Int J Design Analysis Tools Integrated Circuits Syst* 2019;8(1):72-78 [FREE Full text]
26. Müller M, Schneider M, Salathé M, Vayena E. Assessing public opinion on CRISPR-Cas9: combining crowdsourcing and deep learning. *J Med Internet Res* 2020 Aug 31;22(8):e17830 [FREE Full text] [doi: [10.2196/17830](https://doi.org/10.2196/17830)] [Medline: [32865499](https://pubmed.ncbi.nlm.nih.gov/32865499/)]
27. Zhang D, Luo Y, Chen W. [Current development of gene editing]. *Sheng Wu Gong Cheng Xue Bao* 2020 Nov 25;36(11):2345-2356 [FREE Full text] [doi: [10.13345/j.cjb.190557](https://doi.org/10.13345/j.cjb.190557)] [Medline: [33244929](https://pubmed.ncbi.nlm.nih.gov/33244929/)]
28. Zhang Q, Goncalves B. Topical differences between Chinese language Twitter and Sina Weibo. In: Proceedings of the 25th International Conference Companion on World Wide Web. 2016 Presented at: WWW '16: 25th International World Wide Web Conference; Apr 11 - 15, 2016; Montréal Québec Canada. [doi: [10.1145/2872518.2890562](https://doi.org/10.1145/2872518.2890562)]
29. Zhang L, Huang X, Liu T, Li A, Chen Z, Zhu T. Using linguistic features to estimate suicide probability of Chinese microblog users. In: *Human Centered Computing*. Cham: Springer; 2015.
30. Huang C, Xu X, Cai Y, Ge Q, Zeng G, Li X, et al. Mining the characteristics of COVID-19 patients in China: analysis of social media posts. *J Med Internet Res* 2020 May 17;22(5):e19087 [FREE Full text] [doi: [10.2196/19087](https://doi.org/10.2196/19087)] [Medline: [32401210](https://pubmed.ncbi.nlm.nih.gov/32401210/)]
31. Competitive analysis, marketing mix and traffic. Alexa. URL: <https://www.alexa.com/> [accessed 2022-02-20]
32. Li C, Lu Y, Mei Q, Wang D, Pandey S. Click-through prediction for advertising in twitter timeline. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2015 Presented at: KDD '15: The 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; Aug 10 - 13, 2015; Sydney NSW Australia. [doi: [10.1145/2783258.2788582](https://doi.org/10.1145/2783258.2788582)]
33. Phuvipadawat S, Murata T. Breaking news detection and tracking in Twitter. In: Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. 2010 Presented at: 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology; Aug 31 - Sep 3, 2010; Toronto, ON, Canada. [doi: [10.1109/wi-iat.2010.205](https://doi.org/10.1109/wi-iat.2010.205)]
34. Poecze F, Ebster C, Strauss C. Social media metrics and sentiment analysis to evaluate the effectiveness of social media posts. *Procedia Comput Sci* 2018;130:660-666. [doi: [10.1016/j.procs.2018.04.117](https://doi.org/10.1016/j.procs.2018.04.117)]
35. Schwemmer C, Ziewiecki S. Social media sellout: the increasing role of product promotion on YouTube. *Social Media Soc* 2018 Aug 14;4(3):205630511878672. [doi: [10.1177/2056305118786720](https://doi.org/10.1177/2056305118786720)]
36. Dean B. Reddit user and growth stats (updated Oct 2021). BACKLINKO. 2021. URL: <https://backlinko.com/reddit-users> [accessed 2021-10-11]
37. Sattelberg W. The demographics of reddit: who uses the site? alphr. URL: <https://www.alphr.com/demographics-reddit/> [accessed 2022-02-20]
38. Yin F, Zhang B, Zhou J, Liu J. The public opinion analysis scheme and method research on multi-platform. In: Proceedings of the 2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP). 2015 Presented at: 2015 12th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP); Dec 18-20, 2015; Chengdu, China. [doi: [10.1109/iccwamtip.2015.7493960](https://doi.org/10.1109/iccwamtip.2015.7493960)]
39. Roberts M, Stewart B, Tingley D, Airoidi E. The structural topic model and applied social science. NIPS 2013 Workshop. URL: <https://scholar.princeton.edu/files/bstewart/files/stmnips2013.pdf> [accessed 2022-02-20]

40. Chuang J, Manning C, Heer J. Termite: visualization techniques for assessing textual topic models. In: Proceedings of the International Working Conference on Advanced Visual Interfaces. 2012 Presented at: AVI'12: International Working Conference on Advanced Visual Interfaces; May 21 - 25, 2012; Capri Island Italy. [doi: [10.1145/2254556.2254572](https://doi.org/10.1145/2254556.2254572)]
41. Sievert C, Shirley K. LDavis: a method for visualizing and interpreting topics. In: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces. 2014 Presented at: Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces; Jun 27, 2014; Baltimore, Maryland, USA. [doi: [10.3115/v1/W14-3110](https://doi.org/10.3115/v1/W14-3110)]
42. Cohen J. Russian geneticist answers challenges to his plan to make gene-edited babies. *Science*. 2019. URL: <https://tinyurl.com/adphxuzu> [accessed 2022-02-20]
43. Ledford H. CRISPR babies: when will the world be ready? *Nature* 2019 Jun;570(7761):293-296. [doi: [10.1038/d41586-019-01906-z](https://doi.org/10.1038/d41586-019-01906-z)] [Medline: [31217610](https://pubmed.ncbi.nlm.nih.gov/31217610/)]
44. Thompson J. What we risk as humans if we allow gene-edited babies: a philosopher's view. *The conversation*. URL: <https://theconversation.com/what-we-risk-as-humans-if-we-allow-gene-edited-babies-a-philosophers-view-110498> [accessed 2022-02-20]
45. Exclusive: five couples lined up for CRISPR babies to avoid deafness. *New Scientist*. 2019. URL: <https://www.newscientist.com/article/2208777-exclusive-five-couples-lined-up-for-crispr-babies-to-avoid-deafness/> [accessed 2022-02-20]
46. *Dual Use Research of Concern in the Life Sciences*. Washington, D.C., United States: National Academies Press; 2017.
47. Global governance of emerging technology: public backlash and regulatory competition in gene editing. *Square Space*. URL: https://static1.squarespace.com/static/5c54747ad7819e06b909fb1f/t/605b509abb0633544527737e/1616597149243/Genes_March2021.pdf [accessed 2022-02-20]
48. Critchley C, Nicol D, Bruce G, Walshe J, Treleaven T, Tuch B. Predicting public attitudes toward gene editing of germlines: the impact of moral and hereditary concern in human and animal applications. *Front Genet* 2018;9:704 [FREE Full text] [doi: [10.3389/fgene.2018.00704](https://doi.org/10.3389/fgene.2018.00704)] [Medline: [30687386](https://pubmed.ncbi.nlm.nih.gov/30687386/)]
49. Uchiyama M, Nagai A, Muto K. Survey on the perception of germline genome editing among the general public in Japan. *J Hum Genet* 2018 Jun;63(6):745-748 [FREE Full text] [doi: [10.1038/s10038-018-0430-2](https://doi.org/10.1038/s10038-018-0430-2)] [Medline: [29545588](https://pubmed.ncbi.nlm.nih.gov/29545588/)]
50. van Dijke I, van Wely M, Berkman B, Bredenoord AL, Henneman L, Vliegenthart R, et al. Should germline genome editing be allowed? The effect of treatment characteristics on public acceptability. *Hum Reprod* 2021 Jan 25;36(2):465-478 [FREE Full text] [doi: [10.1093/humrep/deaa212](https://doi.org/10.1093/humrep/deaa212)] [Medline: [33242333](https://pubmed.ncbi.nlm.nih.gov/33242333/)]
51. Jedwab A, Vears DF, Tse C, Gyngell C. Genetics experience impacts attitudes towards germline gene editing: a survey of over 1500 members of the public. *J Hum Genet* 2020 Dec 31;65(12):1055-1065. [doi: [10.1038/s10038-020-0810-2](https://doi.org/10.1038/s10038-020-0810-2)] [Medline: [32737393](https://pubmed.ncbi.nlm.nih.gov/32737393/)]
52. Howell EL, Yang S, Beets B, Brossard D, Scheufele DA, Xenos MA. What do we (not) know about global views of human gene editing? Insights and blind spots in the CRISPR era. *CRISPR J* 2020 Jun;3(3):148-155. [doi: [10.1089/crispr.2020.0004](https://doi.org/10.1089/crispr.2020.0004)] [Medline: [33560915](https://pubmed.ncbi.nlm.nih.gov/33560915/)]
53. Calabrese C, Ding J, Millam B, Barnett GA. The uproar over gene-edited babies: a semantic network analysis of CRISPR on Twitter. *Environ Commun* 2019 Dec 19;14(7):954-970. [doi: [10.1080/17524032.2019.1699135](https://doi.org/10.1080/17524032.2019.1699135)]
54. Liu Y, Lapata M. Text summarization with pretrained encoders. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019 Presented at: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); Nov 3-7, 2019; Hong Kong, China. [doi: [10.18653/v1/d19-1387](https://doi.org/10.18653/v1/d19-1387)]
55. Jovanovic D, Van Leeuwen T. Multimodal dialogue on social media. *Soc Semiotics* 2018 Aug 28;28(5):683-699. [doi: [10.1080/10350330.2018.1504732](https://doi.org/10.1080/10350330.2018.1504732)]
56. Lu D, Neves L, Carvalho V, Zhang N, Ji H. Visual attention model for name tagging in multimodal social media. In: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 2018 Presented at: Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers); Jul 15-20, 2018; Melbourne, Australia. [doi: [10.18653/v1/p18-1185](https://doi.org/10.18653/v1/p18-1185)]
57. Schifanella R, de Juan P, Tetreault J, Cao L. Detecting sarcasm in multimodal social platforms. In: Proceedings of the 24th ACM international conference on Multimedia. 2016 Presented at: MM '16: ACM Multimedia Conference; Oct 15 - 19, 2016; Amsterdam The Netherlands. [doi: [10.1145/2964284.2964321](https://doi.org/10.1145/2964284.2964321)]
58. Xiang Z, Du Q, Ma Y, Fan W. A comparative analysis of major online review platforms: implications for social media analytics in hospitality and tourism. *Tourism Manag* 2017 Feb;58:51-65. [doi: [10.1016/j.tourman.2016.10.001](https://doi.org/10.1016/j.tourman.2016.10.001)]
59. Zhan Y, Liu R, Li Q, Leischow SJ, Zeng DD. Identifying topics for e-cigarette user-generated contents: a case study from multiple social media platforms. *J Med Internet Res* 2017 Jan 20;19(1):e24 [FREE Full text] [doi: [10.2196/jmir.5780](https://doi.org/10.2196/jmir.5780)] [Medline: [28108428](https://pubmed.ncbi.nlm.nih.gov/28108428/)]
60. Primack BA, Shensa A, Escobar-Viera CG, Barrett EL, Sidani JE, Colditz JB, et al. Use of multiple social media platforms and symptoms of depression and anxiety: a nationally-representative study among U.S. young adults. *Comput Human Behav* 2017 Apr;69:1-9. [doi: [10.1016/j.chb.2016.11.013](https://doi.org/10.1016/j.chb.2016.11.013)]

61. Ruiz J, Featherstone J, Barnett G. Identifying vaccine hesitant communities on twitter and their geolocations: a network approach. In: Proceedings of the 54th Hawaii International Conference on System Sciences. 2021 Presented at: 54th Hawaii International Conference on System Sciences; Jan 5, 2021; Hawaii. [doi: [10.24251/hicss.2021.480](https://doi.org/10.24251/hicss.2021.480)]
62. Hurlbut JB. Human genome editing: ask whether, not how. *Nature* 2019 Jan;565(7738):135. [doi: [10.1038/d41586-018-07881-1](https://doi.org/10.1038/d41586-018-07881-1)] [Medline: [30602779](https://pubmed.ncbi.nlm.nih.gov/30602779/)]

Abbreviations

API: application programming interface
Cas9: CRISPR-associated protein 9
CRISPR: clustered regularly interspaced palindromic repeats
GGE: germline gene editing
HGE: human gene editing
STM: structural topic modeling

Edited by G Eysenbach; submitted 30.06.21; peer-reviewed by G Barnett, K Kades, A Louren; comments to author 26.08.21; revised version received 19.10.21; accepted 24.01.22; published 11.03.22.

Please cite as:

Ni C, Wan Z, Yan C, Liu Y, Clayton EW, Malin B, Yin Z

The Public Perception of the #GeneEditedBabies Event Across Multiple Social Media Platforms: Observational Study

J Med Internet Res 2022;24(3):e31687

URL: <https://www.jmir.org/2022/3/e31687>

doi: [10.2196/31687](https://doi.org/10.2196/31687)

PMID: [35275077](https://pubmed.ncbi.nlm.nih.gov/35275077/)

©Congning Ni, Zhiyu Wan, Chao Yan, Yongtai Liu, Ellen Wright Clayton, Bradley Malin, Zhijun Yin. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 11.03.2022. 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.

Original Paper

Puff Bars, Tobacco Policy Evasion, and Nicotine Dependence: Content Analysis of Tweets

Kar-Hai Chu^{1*}, PhD; Tina B Hershey^{1*}, JD, MPH; Beth L Hoffman^{1*}, MPH; Riley Wolynn^{1*}; Jason B Colditz^{2*}, PhD; Jaime E Sidani^{1*}, MPH, PhD; Brian A Primack^{3*}, MD, PhD

¹Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA, United States

²School of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

³College of Education and Health Professions, University of Arkansas, Fayetteville, AR, United States

*all authors contributed equally

Corresponding Author:

Kar-Hai Chu, PhD

Graduate School of Public Health

University of Pittsburgh

130 DeSoto Street

Pittsburgh, PA, 15261

United States

Phone: 1 4126922578

Email: chuk@pitt.edu

Abstract

Background: Puff Bars are e-cigarettes that continued marketing flavored products by exploiting the US Food and Drug Administration exemption for disposable devices.

Objective: This study aimed to examine discussions related to Puff Bar on Twitter to identify tobacco regulation and policy themes as well as unanticipated outcomes of regulatory loopholes.

Methods: Of 8519 original tweets related to Puff Bar collected from July 13, 2020, to August 13, 2020, a random 20% subsample (n=2661) was selected for qualitative coding of topics related to nicotine dependence and tobacco policy.

Results: Of the human-coded tweets, 2123 (80.2%) were coded as relevant to Puff Bar as the main topic. Of those tweets, 698 (32.9%) discussed tobacco policy, including flavors (n=320, 45.9%), regulations (n=124, 17.8%), purchases (n=117, 16.8%), and other products (n=110, 15.8%). Approximately 22% (n=480) of the tweets referenced dependence, including lack of access (n=273, 56.9%), appetite suppression (n=59, 12.3%), frequent use (n=47, 9.8%), and self-reported dependence (n=110, 22.9%).

Conclusions: This study adds to the growing evidence base that the US Food and Drug Administration ban of e-cigarette flavors did not reduce interest, but rather shifted the discussion to brands utilizing a loophole that allowed flavored products to continue to be sold in disposable devices. Until comprehensive tobacco policy legislation is developed, new products or loopholes will continue to supply nicotine demand.

(*J Med Internet Res* 2022;24(3):e27894) doi:[10.2196/27894](https://doi.org/10.2196/27894)

KEYWORDS

tobacco; policy; social media; e-cigarette; twitter; mHealth; dependence; addiction; nicotine

Introduction

From 2011 to 2019, current e-cigarette use among US high school students increased from 1.5% to 27.5% [1], prompting the US Surgeon General to declare a youth vaping epidemic. An appealing aspect of e-cigarette use to adolescents was the availability of flavors [2]. Among this population, initial use of flavored e-cigarettes is associated with progression to current e-cigarette use [3].

Driven by the results of the 2019 National Youth Tobacco Survey (NYTS) and other reports of increased use of tobacco products by youth, the US Food and Drug Administration (FDA) took action to address this epidemic in December 2019 by raising the federal minimum age for sale of tobacco products (including e-cigarettes) from 18 to 21 years and by prioritizing enforcement against illegal flavored (eg, fruits) e-cigarettes [4,5]. Likewise, many US states have enacted legislation to restrict flavored e-cigarettes [6]. In an effort to balance

considerations for adult smokers trying to quit cigarettes, the federal flavor ban was focused on cartridge-based products, such as those sold by JUUL (JUUL Labs)—at the time, the device used by a majority of youth who were current e-cigarette users [1].

However, the actions by the FDA may have resulted in unintended consequences. In the case of e-cigarette-related policy, loopholes allowed for disposable devices such as Puff Bar to continue to be sold, even in prohibited flavors. Puff Bars are single-use, disposable, flavored e-cigarette products. The

design and packaging of Puff Bar are similar to those of JUUL (Figure 1). Puff Bar e-cigarettes come in 25 different flavors (eg, strawberry banana). There is evidence that Puff Bar is targeting its products and advertisements to youth. For example, the company produced flavor pods (Puff Krush) that are advertised as an “add-on” for JUUL pods following JUUL’s removal of most flavors from the US market, which were extremely popular with youth [7]. In 2020, the NYTS reported that the use of disposable e-cigarettes among current high school users increased by approximately 1000% from 2019 [8].

Figure 1. Puff Bar and JUUL comparison. Left: Puff Bar device; right: JUUL device.



On July 13, 2020, Puff Bar announced that they were ceasing online sales in the United States [9]. One week later, the FDA announced that it issued warning letters to 10 companies, including Cool Clouds Distribution, Inc (Puff Bar’s parent company), to remove their products from the market, citing their introduction after the 2016 deeming rule bringing all tobacco products under the authority of the FDA [10]. Puff Bar was also cited for marketing their product as a modified risk tobacco product without FDA approval [10]. However, evidence suggests that Puff Bar sales were continuing despite the FDA warnings. For example, at the time this article was written, the Puff Bar website still had a “store locator” function [11] listing retailers across the US.

The continued sales of flavored products in disposable devices may be contributing to youth use of products containing high levels of nicotine. In contrast to contemporary products that contain freebase nicotine, Puff Bar e-cigarettes contain nicotine salt formulations (similar to JUUL) that deliver nicotine in a quickly metabolized and palatable manner, with nicotine concentrations as high as 5% [12,13]. Research indicates that

users of JUUL’s higher nicotine level products (ie, 5%) experience symptoms of dependence and acute nicotine effects [14]. Likewise, nicotine dependence in past-month adolescent e-cigarette users is significantly associated with increased nicotine concentrations [15]. Thus, despite the intended goal of reducing youth tobacco use through legislative and policy activities, unintended loopholes allowed youth to access the same products, just in a different form.

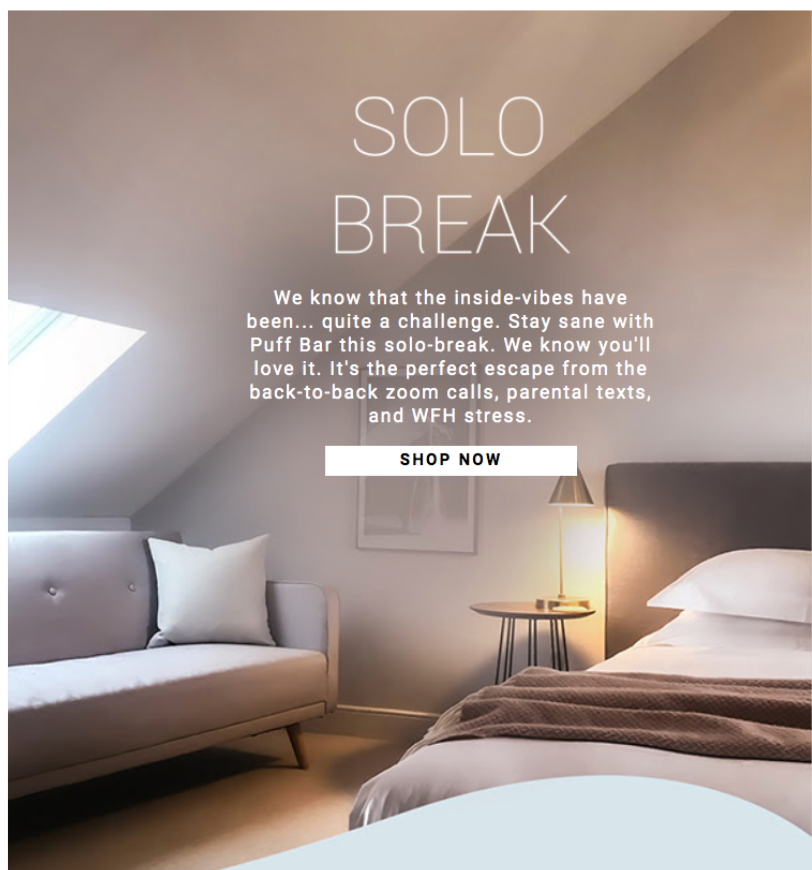
Research has consistently found that e-cigarette discussions on social media platforms can quickly diffuse marketing messages, activate brand awareness, and reach large numbers of adolescents [16-18]. An analysis of videos on TikTok, a social media platform particularly popular with younger populations [19], found that the 10 most popular videos depicting Puff Bar were viewed between 2.8 and 42.4 million times and that 2 of the videos depicted clear youth-related content (eg, an underage youth purchasing a Puff Bar) [20]. Additionally, during the COVID-19 pandemic social distancing restrictions, Puff Bar released an advertisement picturing a bedroom and suggested that their products would allow for escape from “back-to-back

zoom calls, parental texts, and WFH [work from home] stress” (Figure 2) [21].

With growing risk of adolescent use of disposable e-cigarette devices, studying Puff Bar on social media may offer insight into topics of discussion and trends that are emerging [22]. Twitter has become a valuable source of publicly observable data for public health practitioners to better understand attitudes toward e-cigarette use, advertisements targeted to youth, and discussions of tobacco regulations [23-26]. Pew Research has also found that Twitter users tend to be younger than the overall

US adult population [27]. Specifically, Puff Bar-related discussions on Twitter may give insight into how these products are being used as an alternative to products that fall under federal and state restrictions. Evaluating the potential impact of legal actions related to e-cigarettes will provide important information to the public health community. Thus, this study sought to examine Puff Bar-related discussions on Twitter to identify themes related to tobacco policy as well as unanticipated outcomes of legal and regulatory loopholes, such as effects from continued use of flavored, high nicotine concentration salt formulations, for use in future research.

Figure 2. Puff Bar advertisement during global pandemic.



Methods

Sample Selection

We used the Real-time Inveillance of Twitter Health Messages software framework to collect Twitter posts (ie, tweets) containing the terms “puffbar” or “puff bar” for 1 month from July 13, 2020, to August 13, 2020 [28]. The Real-time Inveillance of Twitter Health Messages allows for real time collection of all publicly available tweets matching a specified set of keywords through Twitter’s filtered data stream application programming interface. This start date aligned with the date Puff Bar announced ceasing online sales in the United States [9]. This resulted in 13,304 tweets, of which 4785 (36%) were retweets and 8519 (64%) were original tweets.

We obtained a 20% random subsample of original tweets (n=2661) for human coding; this process has previously

demonstrated to be both feasible for human qualitative coding and generalizable to the full data set [28].

Ethical Approval

The University of Pittsburgh Institutional Review Board (IRB) determined that the proposed activity is not research involving human subjects as defined by DHHS and FDA regulations (STUDY19080214).

Codebook Development

We developed a codebook based on a hybrid process that included consideration of our research question, prior analysis of e-cigarette discourse on Twitter [14,29], tobacco policy, and an inductive analysis of 100 relevant tweets that were not included in the final sample. First, we included a code for relevance to the research topic of Puff Bar-related discussions (relevant). Tweets that did not contain the disposable e-cigarette Puff Bar as the main topic (eg, “I go to my car to get my puff

bar and somebody left me a rose on my car. I can't stop smiling!") were deemed not relevant.

Relevant tweets were coded for topics informed by 2 areas of research, which were nicotine dependence and tobacco policy. Topics for nicotine dependence included language that suggests dependence on Puff Bar and signs of nicotine addiction or withdrawal related to Puff Bar use. These categories were informed by prior research that examined similar content posted about JUUL [14]. Topics for tobacco policy included whether

posted content were commercially marketing Puff Bar or a business selling Puff Bar, references to purchasing a Puff Bar, references to underage use of Puff Bar, regulations of Puff Bar, price of Puff Bar, references to Puff Bar flavor, and references to other e-cigarette products. The codebook was validated through analysis of 100 relevant tweets that were not included in the final sample by 2 experienced Twitter coders. Following this, the final codebook was codified, presenting clear definitions and examples for each code (Table 1).

Table 1. Definitions for categorical codes and example tweets. Examples are paraphrased.

Code and subcode	Definition	Examples
Tobacco policy	Puff Bar in relationship to laws or regulations that could affect use of Puff Bar	
Commercial	Marketing for Puff Bar or shops selling Puff Bar	<ul style="list-style-type: none"> • Check out what our customers are saying about Puff Bar Disposable Pod device!
Black market	Reference to illegal purchase of a Puff Bar or buying a knockoff	<ul style="list-style-type: none"> • Yo the puff bar black market is wild • I hope the govt knows they created both the fake puff bar prob & the fake the oil problems all by themselves
Buying	Concretely obtaining or trying to obtain Puff Bar	<ul style="list-style-type: none"> • Can I get a puffbar on Instacart? • Got a free puff bar at the gas station :)
Underage use	Underage use (ie, under 21) of Puff Bar	<ul style="list-style-type: none"> • I wish I could explain to you guys how flabbergasted I am to have just met a 5 year old child 2/ a puff bar...he talked about disposable vapes for 5 minutes & rated various flavors • This little boy really asked if I could buy him a puff bar and when I did, he went goat and then says I can't get it anymore I'm sorry
Regulations	Regulations on Puff Bar (cannot buy, cannot access, etc)	<ul style="list-style-type: none"> • Why do I gotta be 21 to buy myself a puff bar • FDA calls for removal of fruity and disposable Puff Bar vapes devices
Price	Price of a Puff Bar	<ul style="list-style-type: none"> • Anyways does anyone wanna paypal me 16 dollar so I can buy a fucking puff bar • Maaan I shouldn't have hit ur puff bar bc now my ass is spending \$15-20 every other week on nic
Flavors	Flavors of Puff Bar	<ul style="list-style-type: none"> • Strawberry banana puff bar is just vaping a gogurt • Peach ice and lychee seemed to be the favorite/most popular flavors
Other products	Other vaping products, including Puff XXL or JUUL	<ul style="list-style-type: none"> • Checkout the newest Puff XXL 1600 disposable device • Why did nobody tell me that air bars get me more puffs than a puff bar plus they r cheaper
References to COVID-19	Use of Puff Bar during the pandemic	<ul style="list-style-type: none"> • I hate when my puff bar starts spitting at me. C'mon bitch were in a pandemic!
Dependence	Puff Bar in association with words that specify dependence on Puff Bar	
Puff bar as a meal	Puff Bar as a meal or in place of food, using terms such as meal, breakfast, lunch, or dinner	<ul style="list-style-type: none"> • My power meal today: one puff bar plus an energy drink • Had a whole lychee puff bar as all of my meals today. I'm thriving babeee
Losing access	Not being able to use Puff Bar to an external factor such as losing it or the battery dying	<ul style="list-style-type: none"> • On 4 hour drive and my puff bar completely died • I lost my puff bar for an hour and found it in my bra. Whoops
Self-report	Self-report of being addicted or dependent on Puff Bar	<ul style="list-style-type: none"> • Just tried a puff bar and I'm definitely addicted • Bought myself a puff bar. Time to bring back my nicotine addiction
Does not last	Using up Puff Bar quickly	<ul style="list-style-type: none"> • I've never been able to make a puff bar last more than 48 hours • Another day. Time for another puff bar
Acute nicotine effects	Puff Bar in association with words that specify acute nicotine effects	<ul style="list-style-type: none"> • My puff bar got me buzzing like a bee • That puffbar feeling - lightheadedness
Quitting or withdrawal	Quitting Puff Bar or experiencing signs of nicotine withdrawal from lack of Puff Bar	<ul style="list-style-type: none"> • I have a problem, so after this puff bar runs out, I'm NOT buying another one • Only positive of being at my parents house is that i finally quit my puffbar quarantine habit

Coding Procedures

The tweets were coded by 2 independent individuals, with adjudication of disagreements by a supervising researcher. The coders were provided with the tweet text and a URL link to each tweet. During the coding process, all relevant tweets that were publicly available at the time of coding were viewed on Twitter so that visuals in the tweet (eg, images, videos, and emoji) could also be assessed. The text from unavailable tweets was still included in the coding and thematic analysis to preserve comprehensiveness of the original data. The codes were not mutually exclusive.

We calculated interrater reliability using the Cohen κ , and disagreements were adjudicated between the 2 coders. In instances where the coders could not reach consensus, the lead author had final determination. After 4 rounds of independent double coding (100 tweets each round), interrater reliability was considered good to excellent (Cohen $\kappa=0.71-1.00$) for all categories [30]. The remaining tweets were split between the 2 coders for independent coding.

Analysis

We calculated descriptive statistics for each coding category and used a thematic content analysis approach to analyze qualitative data [31]. To conduct the content analysis, the coders wrote annotations and memos throughout the coding process and highlighted specific words or phrases within tweets that exemplified the themes. The coders then met with supervising researchers to synthesize themes with salient examples from the observed data.

This study was approved by the University of Pittsburgh Institutional Review Board. To protect tweeters from identification, all examples provided in the text and tables are paraphrased versions of original tweets.

Results

Of the random sample of tweets ($n=2661$), 80.2% ($n=2123$) were coded as relevant to the research question. Of these relevant tweets, 698 (32.9%) tweets discussed topics relevant to tobacco policy (Table 2). In table 2, the percentages for subcodes are presented as the percent of tweets within total tweets for that code.

Table 2. Frequencies of coding categories for relevant tweets ($n=2123$). Categories are not mutually exclusive; therefore, proportions will not always add up to 100%.

Code and subcode	Frequency, n (%)
Tobacco policy	698 (32.9)
Flavors	320 (45.9)
Regulations	124 (17.8)
Buying	117 (16.8)
Other products	110 (15.8)
Black market	31 (4.4)
Price	36 (5.2)
Underage use	24 (3.4)
Commercial	20 (2.9)
Dependence	480 (22.6)
Losing access	273 (56.9)
Self-report	110 (22.9)
Puff bar as a meal	59 (12.3)
Does not last	47 (9.8)
Quitting or withdrawal	52 (2.4)
Acute nicotine effects	50 (2.4)
References to COVID-19	11 (0.5)

The most frequent themes relevant to tobacco policy were references to Puff Bar flavors ($n=320$, 45.9%; eg, “I bought a peach ice puff bar and it is so yummy”). There was a similar prevalence of references to buying Puff Bar ($n=117$, 16.8%; eg, “setting my alarm to go the puff bar store tomorrow morning”) and other products ($n=110$, 15.8%; eg, “a puff bar only lasts me a day so switched to viva. Last 3+ days”).

Approximately 22% ($n=480$) of the tweets referenced dependence in the context of Puff Bar. A majority of these tweets referenced losing access to Puff Bar ($n=273$, 56.9%; eg, “on a long drive and my puff bar died. Send help. Im so upset”). Another theme was users tweeting about their Puff Bar lasting less than 48 hours due to frequent use ($n=47$, 9.8%; eg, “I wonder why I can’t breathe yet I go through a puff bar in 2 days”). Additionally, approximately a quarter ($n=110$, 22.9%)

of tweets referencing dependence involved the user self-reporting dependence on Puff Bar (eg, “I need a puff bar so bad, but I will stay strong. I want to get rid of this mf addiction”).

Users also reported acute nicotine effects, such as feeling a buzz or high when using Puff Bar (n=50, 2.4%; eg, “this puff bar lightheadedness feels so good”). Users reported other symptoms related to nicotine, such as headaches (eg, “is my headache a sign of coronavirus or my puff bar addiction?”) and upset stomach (eg, “I hit my puff bar until my stomach hurts everyday haha”).

Discussion

Principal Findings

This study examined Puff Bar-related discussions on Twitter to identify themes related to tobacco policy and dependence. Despite federal regulations and FDA warnings against Puff Bar, the results of this study suggest that the purchasing and use of Puff Bar products are still being discussed on social media. Discussions of Puff Bar flavors were prevalent, accounting for a plurality (45.9%) of tweets classified as relevant to tobacco policy; these products are circumventing a federal ban designed to protect youth from the appeal of nicotine products. Similarly, the second most common theme was focused on regulations (n=124, 17.8%). As a result, the outcome of the FDA ban of e-cigarette flavors did not reduce demand, but rather shifted more attention to different e-cigarette brands that utilize a loophole for disposable devices.

We found similar dependence and acute nicotine effect themes as prior research on JUUL products and cigarettes, such as self-report of dependence, frustration over losing access to the device, compulsive use, and self-reports of physical effects of nicotine exposure including suppression of appetite [14,32]. Again, this suggests that despite legislative and policy activities to reduce youth access to and use of flavored, high nicotine-containing and dependence-forming products, they continued to be available due to exploitation of loopholes. The fact that high school students’ use of disposable e-cigarettes such as Puff Bar increased by 1000% from 2019 to 2020 is evidence that these loopholes can contribute to continued use of e-cigarettes by youth.

Due to new federal tobacco policies, discussions about Puff Bar offer a view into topics relevant to tobacco policy including e-cigarette regulations and how they might affect access to Puff Bar (17.9%) and buying Puff Bar (16.8%). The results suggest that while the legality of Puff Bar sales might be recognized and discussed by Twitter users, people continue to purchase them even though the products are banned in the United States. This is especially alarming considering that commercial content (ie, posts by retailers) were very low (2.9% of tobacco policy tweets, 0.9% of all relevant tweets) when compared with other e-cigarette research [33,34].

The results of this study align with data from the NYTS that demonstrate a shift to disposable e-cigarettes following FDA tobacco policies. While the federal flavor ban was enacted to curb youth e-cigarette use, it is unfortunate that exemptions

created loopholes to allow continued access to flavored products. While outside the scope of this study, future research should consider identifying what types of accounts are posting this content. Specifically, understanding whether commercial tweets are being posted by Puff Bar users, commercial vendors, news organizations, or others could help to better inform counter messaging or preventive measures. Additionally, it would be valuable for researchers to be guided by a framework that can address both legal and health concerns. For example, legal epidemiology, the scientific study and deployment of law as a factor in the cause, distribution, and prevention of disease and injury in a population, might inform stronger and more comprehensive policies [35]. Legal epidemiology involves the three following components: (1) legal prevention and control—the study and application of laws and legal practices as interventions to prevent disease and injury and as enablers of effective public health administration; (2) legal etiology—the study of laws and legal practices as causes of disease and injury; and (3) policy surveillance—the ongoing, systematic collection, analysis, and dissemination of information about laws and other policies of importance to health [36]. In combination with continued real time infoveillance of e-cigarette discussions, this framework could help better understand how tobacco policy (eg, age or flavor restrictions) and implementation can affect health outcomes.

Puff Bar acknowledged that their products were excluded from the FDA regulation banning most flavors. On February 21, 2020, Puff Bar posted on their official blog a lament about federal regulations “determined to eliminate vaping as a whole” stating that “disposable devices like Puff Bars and e-liquid used in refillable tank systems can still carry flavors like fruits and desserts” [37]. These, and other similar arguments, have been used by JUUL in order to continue promoting their products, particularly through advertisements targeted to adolescents [38]. Puff Bars—which have now resumed sales—have become the latest device to replace previously banned tobacco products. Until we are able to develop comprehensive tobacco policy, new products that capitalize on policy loopholes will continue to supply the demand for nicotine, particularly by adolescents through targeted advertisements or flavored products. We suggest that a framework such as legal epidemiology could offer a unique lens to understand the confluence of tobacco policy and health outcomes, helping inform policy makers to better understand and strengthen the practical implications of tobacco policies.

Limitations

The results of this study should be considered in the context of the following limitations. Twitter users are not representative of the general population, although it is frequently used by adolescents and young adults [27], a population that also frequently uses e-cigarettes. While interpretation of tweets using qualitative analysis can be subjective, we minimized subjectivity by using a systematic coding procedure and the use of experienced Twitter coders; nonetheless, these tweets are discussion of purchasing, rather than the act itself. We also did not use location data; thus, there is the possibility that some tweets are by non-US users. Finally, our results are constrained by the keywords and time period used in our search parameters.

Future research could expand these parameters to widen the scope of the investigation.

Contribution to Literature

Here, we summarize the key findings of our research: (1) laws and regulations around e-cigarettes are rapidly changing in response to increased concern about the use of the products by youth, and prior research suggests the focus of flavor bans on devices such as JUUL may have created a policy loophole that was filled by disposable devices such as Puff Bar; (2) however, it is not yet known how Puff Bar is being used as an alternative to traditional e-cigarettes that fall under federal and state restrictions; (3) our analysis of tweets related to Puff Bar suggests that the FDA ban of e-cigarette flavors did not reduce interest, but rather shifted the discussion to brands utilizing a loophole for disposable devices and suggests the importance of

using a framework such as legal epidemiology when researching and evaluating tobacco policy.

Conclusion

This study found similar dependence and acute nicotine effect themes in Puff Bar-related discussions on Twitter compared to prior research on JUUL and cigarettes [14]. We also found that discussions about Puff Bar on Twitter provided insight into topics relevant to tobacco policy, including flavors, e-cigarette regulations, and purchasing Puff Bar. Our results, in conjunction with evidence by the NYTS and other data sources, suggest that the FDA ban of e-cigarette flavors did not reduce demand, but rather shifted the supply to brands utilizing a loophole for disposable devices. Until comprehensive tobacco policy legislation is developed, new products or loopholes will continue to supply nicotine demand.

Acknowledgments

We would like to thank Erica Barret for coding of tweets and Michelle Woods for editorial assistance. This work is supported by grant R01CA225773 from the National Cancer Institute. Technical infrastructure was supported through National Science Foundation award numbers ACI-1548562 and 1445606, at the Pittsburgh Supercomputing Center. The funding agencies had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

References

1. Cullen K, Gentzke A, Sawdey M, Chang J, Anic G, Wang T, et al. e-Cigarette Use Among Youth in the United States, 2019. *JAMA* 2019 Dec 03;322(21):2095-2103 [FREE Full text] [doi: [10.1001/jama.2019.18387](https://doi.org/10.1001/jama.2019.18387)] [Medline: [31688912](https://pubmed.ncbi.nlm.nih.gov/31688912/)]
2. Leventhal A, Miech R, Barrington-Trimis J, Johnston L, O'Malley PM, Patrick M. Flavors of e-Cigarettes Used by Youths in the United States. *JAMA* 2019 Dec 03;322(21):2132-2134 [FREE Full text] [doi: [10.1001/jama.2019.17968](https://doi.org/10.1001/jama.2019.17968)] [Medline: [31688891](https://pubmed.ncbi.nlm.nih.gov/31688891/)]
3. Audrain-McGovern J, Rodriguez D, Pianin S, Alexander E. Initial e-cigarette flavoring and nicotine exposure and e-cigarette uptake among adolescents. *Drug Alcohol Depend* 2019 Sep 01;202:149-155 [FREE Full text] [doi: [10.1016/j.drugalcdep.2019.04.037](https://doi.org/10.1016/j.drugalcdep.2019.04.037)] [Medline: [31351341](https://pubmed.ncbi.nlm.nih.gov/31351341/)]
4. FDA finalizes enforcement policy on unauthorized flavored cartridge-based e-cigarettes that appeal to children, including fruit and mint. US Food and Drug Administration. 2020. URL: <https://www.fda.gov/news-events/press-announcements/fda-finalizes-enforcement-policy-unauthorized-flavored-cartridge-based-e-cigarettes-appeal-children> [accessed 2021-03-23]
5. Newly signed legislation raises federal minimum age of sale of tobacco products to 21. US Food and Drug Administration. 2019. URL: <https://www.fda.gov/tobacco-products/ctp-newsroom/newly-signed-legislation-raises-federal-minimum-age-sale-tobacco-products-21> [accessed 2020-11-29]
6. States and localities that have restricted the sale of flavored tobacco products. Campaign for Tobacco-Free Kids. URL: <https://www.tobaccofreekids.org/assets/factsheets/0398.pdf> [accessed 2022-03-10]
7. PUFF Krush add-on pre-filled pods - nicotine free (4 Pack). Eliquidstop. 2020. URL: <https://www.eliquidstop.com/a/s/products/puff-krush-add-on-pods-4-pack> [accessed 2020-09-17]
8. Wang TW, Neff LJ, Park-Lee E, Ren C, Cullen KA, King BA. E-cigarette Use Among Middle and High School Students - United States, 2020. *MMWR Morb Mortal Wkly Rep* 2020 Sep 18;69(37):1310-1312 [FREE Full text] [doi: [10.15585/mmwr.mm6937e1](https://doi.org/10.15585/mmwr.mm6937e1)] [Medline: [32941408](https://pubmed.ncbi.nlm.nih.gov/32941408/)]
9. Wolfe E. Controversial e-Cigarette Company Puff Bar Says It's Suspending U.S. Sales. FairWarning. URL: <https://www.fairwarning.org/2020/07/e-cigarette-company-suspends-sales/> [accessed 2020-09-17]
10. FDA notifies companies, including Puff Bar, to remove flavored disposable e-cigarettes and youth-appealing e-liquids from market for not having required authorization. US Food and Drug Administration. URL: <https://www.fda.gov/news-events/press-announcements/fda-notifies-companies-including-puff-bar-remove-flavored-disposable-e-cigarettes-and-youth> [accessed 2020-09-13]
11. Store Locator - Puff Bar Disposable. Puff Bar. URL: <https://puffbar.com/pages/store-locator> [accessed 2022-03-10]

12. What are Puff Bars? Truth Initiative. 2020. URL: https://truthinitiative.org/research-resources/emerging-tobacco-products/what-are-puff-bars?cid=paidsearch_google_brands_puffbar [accessed 2020-09-17]
13. Puff Bar disposable device. Puff Bar. URL: <https://puffbar.com/collections/puff-bar-disposable> [accessed 2020-09-17]
14. Sidani J, Colditz J, Barrett E, Shensa A, Chu K, James A, et al. I wake up and hit the JUUL: Analyzing Twitter for JUUL nicotine effects and dependence. *Drug Alcohol Depend* 2019 Nov 01;204:107500 [FREE Full text] [doi: [10.1016/j.drugalcdep.2019.06.005](https://doi.org/10.1016/j.drugalcdep.2019.06.005)] [Medline: [31499242](https://pubmed.ncbi.nlm.nih.gov/31499242/)]
15. Morean ME, Krishnan-Sarin S, S O'Malley S. Assessing nicotine dependence in adolescent E-cigarette users: The 4-item Patient-Reported Outcomes Measurement Information System (PROMIS) Nicotine Dependence Item Bank for electronic cigarettes. *Drug Alcohol Depend* 2018 Jul 01;188:60-63 [FREE Full text] [doi: [10.1016/j.drugalcdep.2018.03.029](https://doi.org/10.1016/j.drugalcdep.2018.03.029)] [Medline: [29753155](https://pubmed.ncbi.nlm.nih.gov/29753155/)]
16. Chu K, Colditz J, Primack B, Shensa A, Allem J, Miller E, et al. JUUL: Spreading Online and Offline. *J Adolesc Health* 2018 Nov;63(5):582-586 [FREE Full text] [doi: [10.1016/j.jadohealth.2018.08.002](https://doi.org/10.1016/j.jadohealth.2018.08.002)] [Medline: [30348280](https://pubmed.ncbi.nlm.nih.gov/30348280/)]
17. Chu K, Colditz J, Primack B, Shensa A, Allem J, Miller E, et al. 150. JUULing On Instagram. *Journal of Adolescent Health* 2019 Feb;64(2):S77. [doi: [10.1016/j.jadohealth.2018.10.166](https://doi.org/10.1016/j.jadohealth.2018.10.166)]
18. Willett JG, Bennett M, Hair EC, Xiao H, Greenberg MS, Harvey E, et al. Recognition, use and perceptions of JUUL among youth and young adults. *Tob Control* 2019 Jan;28(1):115-116. [doi: [10.1136/tobaccocontrol-2018-054273](https://doi.org/10.1136/tobaccocontrol-2018-054273)] [Medline: [29669749](https://pubmed.ncbi.nlm.nih.gov/29669749/)]
19. Roumeliotis G, Yang Y, Wang E, Alper A. Exclusive: U.S. opens national security investigation into TikTok - sources. Reuters. 2019. URL: <https://www.reuters.com/article/us-tiktok-cfius-exclusive-idUSKBN1XB4IL> [accessed 2020-09-17]
20. Tan ASL, Weinreich E. #PuffBar: how do top videos on TikTok portray Puff Bars? *Tob Control* 2021 Nov 15;30(6):712-713. [doi: [10.1136/tobaccocontrol-2020-055970](https://doi.org/10.1136/tobaccocontrol-2020-055970)] [Medline: [32934090](https://pubmed.ncbi.nlm.nih.gov/32934090/)]
21. Kaplan S. Lawmakers say Puff Bar used pandemic to market to teens. *The New York Times*. 2020. URL: <https://www.nytimes.com/2020/06/02/health/puff-bar-teens.html> [accessed 2020-09-17]
22. Allem J, Dormanesh A, Majmundar A, Rivera V, Chu M, Unger JB, et al. Leading Topics in Twitter Discourse on JUUL and Puff Bar Products: Content Analysis. *J Med Internet Res* 2021 Jul 19;23(7):e26510 [FREE Full text] [doi: [10.2196/26510](https://doi.org/10.2196/26510)] [Medline: [34279236](https://pubmed.ncbi.nlm.nih.gov/34279236/)]
23. Allem J, Dharmapuri L, Unger JB, Cruz TB. Characterizing JUUL-related posts on Twitter. *Drug Alcohol Depend* 2018 Sep 01;190:1-5 [FREE Full text] [doi: [10.1016/j.drugalcdep.2018.05.018](https://doi.org/10.1016/j.drugalcdep.2018.05.018)] [Medline: [29958115](https://pubmed.ncbi.nlm.nih.gov/29958115/)]
24. Colditz JB, Welling J, Smith NA, James AE, Primack BA. World Vaping Day: Contextualizing Vaping Culture in Online Social Media Using a Mixed Methods Approach. *Journal of Mixed Methods Research* 2017 Apr 09;13(2):196-215. [doi: [10.1177/1558689817702753](https://doi.org/10.1177/1558689817702753)]
25. Sowles SJ, Krauss MJ, Connolly S, Cavazos-Rehg PA. A Content Analysis of Vaping Advertisements on Twitter, November 2014. *Prev Chronic Dis* 2016 Sep 29;13:E139 [FREE Full text] [doi: [10.5888/pcd13.160274](https://doi.org/10.5888/pcd13.160274)] [Medline: [27685432](https://pubmed.ncbi.nlm.nih.gov/27685432/)]
26. Lazard A, Wilcox G, Tuttle H, Glowacki E, Pikowski J. Public reactions to e-cigarette regulations on Twitter: a text mining analysis. *Tob Control* 2017 Dec;26(e2):e112-e116 [FREE Full text] [doi: [10.1136/tobaccocontrol-2016-053295](https://doi.org/10.1136/tobaccocontrol-2016-053295)] [Medline: [28341768](https://pubmed.ncbi.nlm.nih.gov/28341768/)]
27. Smith A, Anderson M. Social media use in 2018. Pew Research Center. 2018. URL: <https://www.pewinternet.org/2018/03/01/social-media-use-in-2018> [accessed 2020-01-24]
28. Colditz JB, Chu K, Emery SL, Larkin CR, James AE, Welling J, et al. Toward Real-Time Inveillance of Twitter Health Messages. *Am J Public Health* 2018 Aug;108(8):1009-1014. [doi: [10.2105/ajph.2018.304497](https://doi.org/10.2105/ajph.2018.304497)]
29. Sidani JE, Colditz JB, Barrett EL, Chu K, James AE, Primack BA. JUUL on Twitter: Analyzing Tweets About Use of a New Nicotine Delivery System. *J Sch Health* 2020 Feb;90(2):135-142 [FREE Full text] [doi: [10.1111/josh.12858](https://doi.org/10.1111/josh.12858)] [Medline: [31828791](https://pubmed.ncbi.nlm.nih.gov/31828791/)]
30. Cohen J. *Statistical power analysis for the behavioral sciences*, 2nd edition. Hillsdale, NJ, US: Lawrence Erlbaum Associates; 1988.
31. Braun V, Clarke V. Using thematic analysis in psychology. *Qualitative Research in Psychology* 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
32. DiFranza JR, Rigotti NA, McNeill AD, Ockene JK, Savageau JA, St Cyr D, et al. Initial symptoms of nicotine dependence in adolescents. *Tob Control* 2000 Sep;9(3):313-319 [FREE Full text] [doi: [10.1136/tc.9.3.313](https://doi.org/10.1136/tc.9.3.313)] [Medline: [10982576](https://pubmed.ncbi.nlm.nih.gov/10982576/)]
33. McCausland K, Maycock B, Leaver T, Wolf K, Freeman B, Jancey J. E-Cigarette Advocates on Twitter: Content Analysis of Vaping-Related Tweets. *JMIR Public Health Surveill* 2020 Oct 14;6(4):e17543 [FREE Full text] [doi: [10.2196/17543](https://doi.org/10.2196/17543)] [Medline: [33052130](https://pubmed.ncbi.nlm.nih.gov/33052130/)]
34. Visweswaran S, Colditz JB, O'Halloran P, Han N, Taneja SB, Welling J, et al. Machine Learning Classifiers for Twitter Surveillance of Vaping: Comparative Machine Learning Study. *J Med Internet Res* 2020 Aug 12;22(8):e17478 [FREE Full text] [doi: [10.2196/17478](https://doi.org/10.2196/17478)] [Medline: [32784184](https://pubmed.ncbi.nlm.nih.gov/32784184/)]
35. Ramanathan T, Hulkower R, Holbrook J, Penn M. Legal Epidemiology: The Science of Law. *J Law Med Ethics* 2017 Mar 01;45(1_suppl):69-72 [FREE Full text] [doi: [10.1177/1073110517703329](https://doi.org/10.1177/1073110517703329)] [Medline: [28661299](https://pubmed.ncbi.nlm.nih.gov/28661299/)]

36. Burris S, Ashe M, Levin D, Penn M, Larkin M. A Transdisciplinary Approach to Public Health Law: The Emerging Practice of Legal Epidemiology. *Annu Rev Public Health* 2016 Mar 18;37(1):135-148 [[FREE Full text](#)] [doi: [10.1146/annurev-publhealth-032315-021841](https://doi.org/10.1146/annurev-publhealth-032315-021841)] [Medline: [26667606](https://pubmed.ncbi.nlm.nih.gov/26667606/)]
37. How Long Does a Puff Bar Last? Puff Bar. URL: <https://puffbar.com/blogs/vape-news> [accessed 2022-03-10]
38. FDA warns JUUL Labs for marketing unauthorized modified risk tobacco products, including in outreach to youth. US Food and Drug Administration. 2019. URL: <https://www.fda.gov/news-events/press-announcements/fda-warns-juul-labs-marketing-unauthorized-modified-risk-tobacco-products-including-outreach-youth> [accessed 2021-07-10]

Abbreviations

FDA: Food and Drug Administration

NYTS: National Youth Tobacco Survey

Edited by G Eysenbach; submitted 11.02.21; peer-reviewed by S Emery, C Watts, JP Allem, M Alvarez de Mon; comments to author 03.04.21; revised version received 13.04.21; accepted 03.02.22; published 25.03.22.

Please cite as:

Chu KH, Hershey TB, Hoffman BL, Wolynn R, Colditz JB, Sidani JE, Primack BA

Puff Bars, Tobacco Policy Evasion, and Nicotine Dependence: Content Analysis of Tweets

J Med Internet Res 2022;24(3):e27894

URL: <https://www.jmir.org/2022/3/e27894>

doi: [10.2196/27894](https://doi.org/10.2196/27894)

PMID: [35333188](https://pubmed.ncbi.nlm.nih.gov/35333188/)

©Kar-Hai Chu, Tina B Hershey, Beth L Hoffman, Riley Wolynn, Jason B Colditz, Jaime E Sidani, Brian A Primack. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 25.03.2022. 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.

Original Paper

Content Analysis of Nicotine Poisoning (Nic Sick) Videos on TikTok: Retrospective Observational Infodemiology Study

Vidya Purushothaman^{1,2}, MAS, MPH, MBBS; Tiana McMann^{1,3,4}, MA; Matthew Nali^{1,2,4}, BA; Zhuoran Li^{1,3,4}, MS; Raphael Cuomo^{1,2}, MPH, PhD; Tim K Mackey^{1,3,4}, MAS, PhD

¹Global Health Policy and Data Institute, San Diego, CA, United States

²Department of Anesthesiology and Infectious Diseases and Global Public Health, University of California San Diego, La Jolla, CA, United States

³Department of Anthropology, University of California San Diego, La Jolla, CA, United States

⁴S-3 Research, San Diego, CA, United States

Corresponding Author:

Tim K Mackey, MAS, PhD

Department of Anthropology

University of California San Diego

9500 Gilman Drive

Postal Code: 0505

La Jolla, CA, 92093

United States

Phone: 1 9514914161

Email: tmackey@ucsd.edu

Abstract

Background: TikTok is a microvideo social media platform currently experiencing rapid growth and with 60% of its monthly users between the ages of 16 and 24 years. Increased exposure to e-cigarette content on social media may influence patterns of use, including the risk of overconsumption and possible nicotine poisoning, when users engage in trending challenges online. However, there is limited research assessing the characteristics of nicotine poisoning–related content posted on social media.

Objective: We aimed to assess the characteristics of content on TikTok that is associated with a popular nicotine poisoning–related hashtag.

Methods: We collected TikTok posts associated with the hashtag #nicsick, using a Python programming package (Selenium) and used an inductive coding approach to analyze video content and characteristics of interest. Videos were manually annotated to generate a codebook of the nicotine sickness–related themes. Statistical analysis was used to compare user engagement characteristics and video length in content with and without active nicotine sickness TikTok topics.

Results: A total of 132 TikTok videos associated with the hashtag #nicsick were manually coded, with 52.3% (69/132) identified as discussing firsthand and secondhand reports of suspected nicotine poisoning symptoms and experiences. More than one-third of nicotine poisoning–related content (26/69, 37.68%) portrayed active vaping by users, which included content with vaping behavior such as vaping tricks and overconsumption, and 43% (30/69) of recorded users self-reported experiencing nicotine sickness, poisoning, or adverse events such as vomiting following nicotine consumption. The average follower count of users posting content related to nicotine sickness was significantly higher than that for users posting content unrelated to nicotine sickness ($W=2350.5$, $P=.03$).

Conclusions: TikTok users openly discuss experiences, both firsthand and secondhand, with nicotine adverse events via the #nicsick hashtag including reports of overconsumption resulting in sickness. These study results suggest that there is a need to assess the utility of digital surveillance on emerging social media platforms for vaping adverse events, particularly on sites popular among youth and young adults. As vaping product use–patterns continue to evolve, digital adverse event detection likely represents an important tool to supplement traditional methods of public health surveillance (such as poison control center prevalence numbers).

(*J Med Internet Res* 2022;24(3):e34050) doi:[10.2196/34050](https://doi.org/10.2196/34050)

KEYWORDS

nic sick; vaping; tobacco; social media; TikTok; content analysis; smoking; nicotine; e-cigarette; adverse effects; public health; infodemiology

Introduction

TikTok is a social media platform that has gained millions of followers and now has brand recognition comparable to that of other well-known social media platforms, such as Facebook, YouTube, and Twitter [1]. TikTok users post, share, and rank microvideos or short clips (between 15 seconds to 3 minutes in duration), with the potential for their content to go viral depending on the level of user interactions, shares, hashtags, and subject matter [2]. In 2021, TikTok reported 1.2 billion monthly active users, and 1.5 billion monthly active users are projected for the end of 2022 [1,3]. It was one of the fastest growing social media platforms, with over 2 billion total downloads since its launch [3]. Approximately 60% of monthly TikTok users in the United States are between the ages of 16 and 24 years [4]—a demographic that is at risk for tobacco initiation and uptake [5]. Additionally, the platform has also been identified as a source of pro-tobacco and tobacco product sentiment [6,7].

Uptake of electronic nicotine delivery systems, as well as patterns of use, can be impacted by social interactions that occur in digital environments as a result of exposure to marketing content, influencers, and other pro-nicotine product sentiment shared by users [8,9]. Increased social media use among adolescents can lead to increased willingness and intention to use e-cigarettes, with reported positive and normative perceptions [10]. Concurrent with the rise in popularity and high rate of adoption among youth and young adults on emerging social media platforms such as TikTok, the increased popularity and use of electronic nicotine delivery systems has given rise to a vaping epidemic [11,12]. For example, from 2016 to 2018, there was a 2.9% increase in smoking prevalence among young adults aged 18 to 24 years in the United States, and more than one-quarter of 10th graders reported having used an electronic nicotine delivery system product [13]. Due to this appeal, vape manufacturers and web-based sellers reportedly use TikTok influencers and direct marketing and selling of electronic nicotine delivery system products on the platform, although public health and tobacco control advocates have also attempted to use TikTok for health promotions warning about the risks of vaping [14].

One of the consequences of increased use of electronic nicotine delivery systems has been the rise in nicotine-related adverse events and poisoning, which was highlighted by the 2019 outbreak of lung injury associated with use of e-cigarette or vaping products that led to 2807 hospitalizations or deaths in the United States [15]. Nicotine poisoning refers to the toxic effects of nicotine consumption that are beyond an individuals' tolerance. Poisoning causes nonspecific symptoms such as nausea, vomiting, tremors, and increased heart rate in early phases and severe symptoms such as shock, low blood pressure, and even paralysis in later phases [16]. Complicating this public safety risk, nicotine poisoning and other electronic nicotine delivery systems adverse events have increased [17,18], partially

due to the introduction of new nicotine products, some of which still require market authorization by regulatory agencies [17]. Most nicotine poisoning research and product surveillance is based on data from government consumer safety organizations (eg, US National Electronic Injury Surveillance System) or poison control centers [19], with only a few studies [20] using social media surveillance to detect adverse events. *Nic sick* is a term that has been used to refer to these adverse effects by users on various social media platforms who post content related to nicotine sickness [21] and can be used as a keyword to curate user-generated self-reporting about nicotine sickness topics and lived experiences.

While TikTok is a platform that can promote and share beneficial health-related content and possibly mobilize public health campaigns [22-24], it can also spread harmful information [25], including content that promotes the use and overuse of electronic nicotine delivery systems to its large audience of youth and young adults [18,26,27]. For example, recent published studies [28,29] that examined health information videos on TikTok observed that the overall quality of health information can vary based on the source of content and that users should be selective and cautious when viewing such content. Additionally, the findings of systematic thematic analysis in 2021 [7] showed that a majority of electronic nicotine delivery system-related TikTok videos portrayed electronic nicotine delivery system-use positively. Another study [30] found that disposable electronic nicotine delivery system product content was popular among TikTok users. Therefore, due to the growing popularity of electronic nicotine delivery system content on the platform, in addition to the high burden of nicotine consumption and risk of nicotine adverse events among TikTok youth and young adults [31], we aimed to expand on prior research by examining the specific characteristics of nicotine poisoning-related TikTok content.

Methods**Ethics**

All information used in this study was posted publicly, and the study did not involve any interaction with users. User identifiable information was removed.

Data Collection

The hashtag #nicsick was originally selected for review based on news and media reports on the popularity of the term for users' self-reporting experiences and topics associated with nicotine sickness [21]. The selection of #nicsick as the hashtag for data collection was confirmed by conducting preliminary manual searches with the TikTok in-platform search function for keywords, terms, and hashtags that were used in videos discussing nicotine sickness content. In addition, we conducted preliminary Google Trends search terms analysis to assess if there were any queries or topics related to the *nic sick* terminology set for the study period, which did not reveal any

specific terms for further keyword or hashtag data filtering. TikTok videos with the hashtag #nicsick were collected retrospectively in May 2021 by first using structured searches on the platform without any user log-in credentials, with no personal search history and with cookies disabled. The URLs and metadata (eg, date, time, username, favorites) associated with retrieved TikTok videos using the hashtag #nicsick were collected and saved using an algorithm (Python, version 3.7.0) and an automated web browser and automation tool (Selenium, version 3.141.0; SeleniumHQ). In addition, the videos were also downloaded and saved for content analysis.

Data Analysis

Content analysis of TikTok videos collected was conducted by VP, TM, and MN. The videos were manually annotated (using binary classification—whether the video depicted nicotine poisoning or nicotine sickness or not) and then inductively coded for specific themes that emerged, such as showing active vaping, specific adverse events, or nicotine sickness experiences (Table 1). Furthermore, based on prior research studies [7,32,33] that have conducted content analysis on TikTok videos, we also coded the following metrics and associated metadata: (1) user engagement (views, likes, comments, shares, followers, verification); (2) video characteristics (duration, caption, text

on screen, subtitles, music); and (3) video type (original, duet, stitch). An *original* video is a microvideo uploaded by a TikTok user and is the primary source of user-generated posts on the platform. A *stitch* integrates the user’s own video into another user’s original video by clipping and integrating scenes from the original video which allows the sharing of existing videos in combination with new user-generated content. A *duet* video builds on another user’s original video, with a new video recorded alongside the original as it plays and is mainly used for user reaction videos. VP, TM, and MN coded posts independently and achieved high intercoder reliability ($\kappa=0.96$). For inconsistent results, senior author TKM was consulted.

After open inductive coding based on our coding schema, videos with nicotine poisoning content were grouped into main thematic categories: (1) firsthand and secondhand reporting of suspected nicotine sickness symptoms (eg, nausea, headache, cough, stomachache, and regurgitation); (2) videos displaying actual poisoning-related adverse events; and (3) videos displaying active vaping while discussing nicotine sickness-related content. The 2-tailed independent sample *t* test and Wilcoxon rank-sum test were used to compare the engagement metrics of videos with and without active nicotine sickness-related content. All statistical analyses were conducted using RStudio (version 3.6.1). A *P* value<.05 was considered statistically significant.

Table 1. Coding scheme for videos.

Theme and subtheme	Coding scheme
Nicotine sickness symptom reporting (firsthand or secondhand)	
Yes	<ul style="list-style-type: none"> • Videos of users discussing specific symptoms related to nicotine sickness, such as vomiting, nausea, burning sensation in throat, headache, and fatigue, during or after nicotine use • Users discussing nicotine sickness symptoms of self or friends, family, and neighbors • Users discussing clinic/hospital/emergency room visits for alleged nicotine sickness symptoms • Videos of users seeking suggestions to overcome nicotine sickness symptoms
No	<ul style="list-style-type: none"> • Posting videos on news posts related to nicotine sickness • Videos on nicotine sickness symptoms by users or public health organizations to create awareness about nicotine sickness • Videos unrelated to nicotine sickness using the hashtag #nicsick • Videos with jokes or sarcasm about nicotine sickness
Displaying poisoning-related adverse events	<ul style="list-style-type: none"> • Videos with users displaying symptoms related to nicotine sickness or poisoning
Actively vaping	<ul style="list-style-type: none"> • Users actively vaping while discussing nicotine sickness-related content in the video

Results

A total of 134 TikTok videos with #nicsick were collected, but 2 videos were inaccessible due to user privacy settings and were excluded from analysis. After manual annotation, we confirmed that 52.3% (69/132) of videos included content discussing nicotine sickness, poisoning, or adverse event symptoms, of which 52 were original videos, 16 were stitches, and 1 was a duet. Of these 69 videos, 36.23% (n=25) were posted in the year 2020 and 63.77% (n=44) in 2021; and the earliest was posted on January 12, 2020, and the latest was posted on May 24, 2021. The videos that were unrelated to nicotine sickness—reporting primarily depicted content related to vaping

behavior and other topics, but nevertheless, included the hashtag #nicsick.

Videos with content related to nicotine sickness discussed specific symptoms, such as vomiting, nausea, burning sensation in throat, headache, and fatigue, during or after nicotine use. Apart from posting videos with content on nicotine sickness experiences, users also asked for suggestions on how to overcome nicotine sickness symptoms from other TikTok users who may have previously experienced similar adverse effects following nicotine consumption. More than one-third of the videos with nicotine poisoning content (26/69, 37.68%) portrayed active vaping (including extreme vaping behavior such as vaping tricks, nicotine overconsumption, attempts at performing vaping challenges) and 43.5% (30/69) recorded

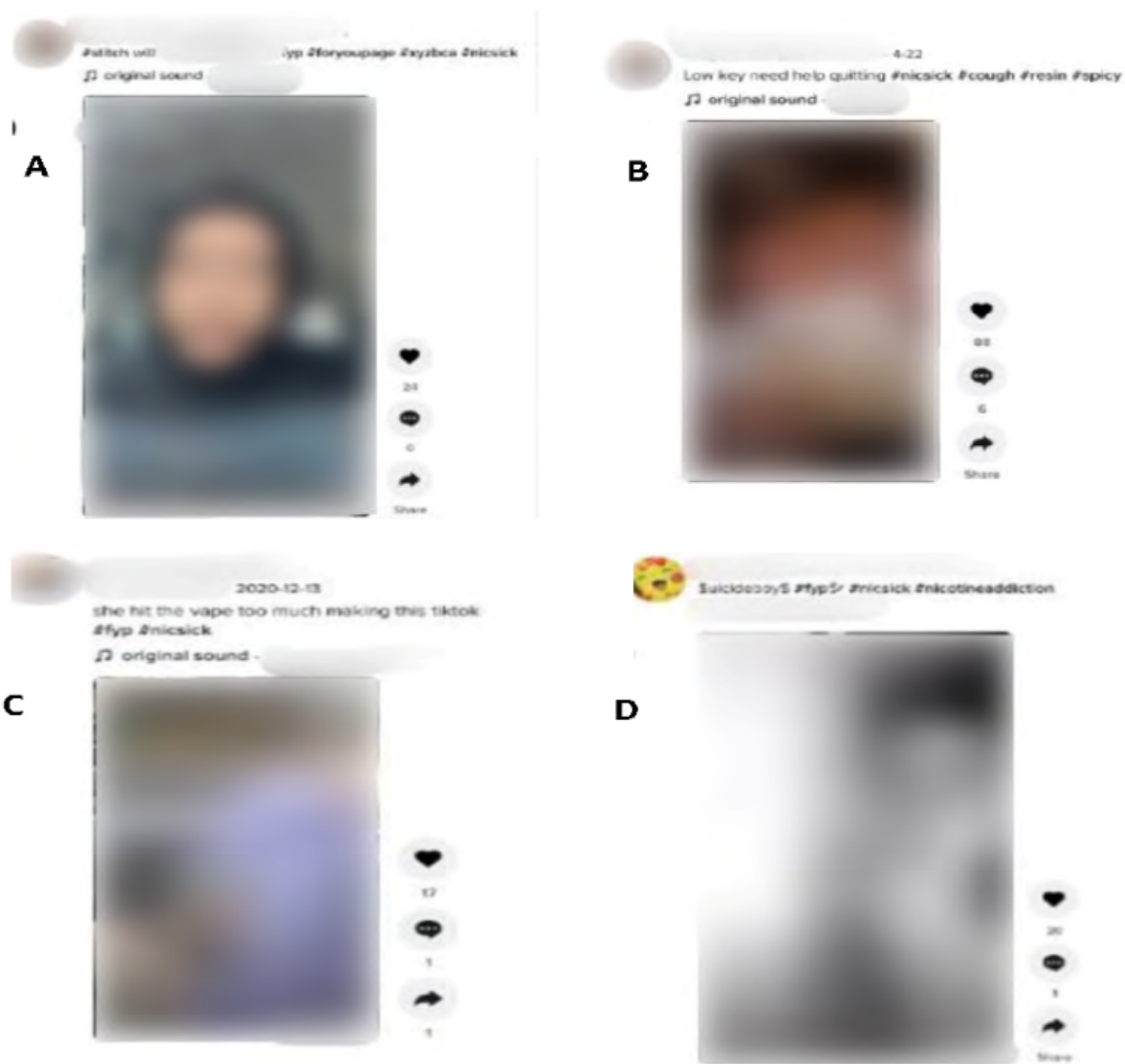
users experiencing nicotine sickness, poisoning, or adverse events, such as vomiting after nicotine consumption. While the first category of videos with only mentioned users experiencing various adverse effects during or after nicotine consumption,

the second category of videos recorded the adverse effects, such as vomiting, nasal discharge, vigorous coughing, directly in the video (Table 2; Figure 1).

Table 2. Descriptions of #nicsick examples (deidentified).

Theme	Definition	Example
Firsthand (self) and secondhand reporting of nicotine sickness symptoms.	A user reporting one or more suspected nicotine sickness symptoms (nausea, cough, regurgitation, etc) currently or in the past.	Figure 1A is a screenshot of a TikTok video in which the user describes their own experience with nicotine sickness
Users displaying adverse events related to nicotine poisoning.	A user exhibiting one or more adverse events (regurgitation, cough, etc) related to nicotine poisoning.	Figure 1B and Figure 1C is a screenshot of a TikTok video in which the user describes their own experience with nicotine sickness Figure 1B shows a user experiencing nasal discharge. Figure 1C shows user vomiting on the side of the road after making a TikTok.
Users actively vaping	A user actively vaping on video.	Figure 1D is a screenshot of a TikTok video in which the user is seen actively vaping using an electronic nicotine delivery systems device.

Figure 1. TikTok video screenshots with signal for nicotine sickness (“nicsick”; user information concealed and images blurred).



A total of 67 TikTok users posted the 69 videos with nicotine sickness content, of which 66 users were unique, while only 1 user (with a follower count of 2487) posted 3 videos, with 2 of them discussing adverse events after nicotine consumption. The

63 videos with content unrelated to nicotine sickness were posted by 60 users, of which 57 user IDs were unique, while 3 users posted 2 videos each. Upon comparing the user IDs, we

found that 3 users had posted signal and nonsignal videos using the hashtag #nicsick.

Videos with nicotine sickness content had 4415 mean views, 548 mean likes, 4 mean shares, 16 mean comments, and a mean user follower count of 19,485, and videos unrelated to nicotine sickness had 52,117 mean views, 1958 mean likes, 31 mean shares, 30 mean comments, and a mean user follower count of 8669. The follower count of users posting content related to nicotine sickness was significantly higher ($W=2350.5$; $P=.03$) than that of users posting content unrelated to nicotine sickness. The mean differences for views ($W=1688.5$; $P=.45$), likes ($W=1780$; $P=.56$), shares ($W=1733$; $P=.36$), and comments ($W=1811.5$; $P=.66$) between videos with nicotine sickness content and those without were not statistically significant.

The mean duration of videos with nicotine sickness content was 23.68 seconds (range 4–62 seconds) and the mean duration of videos displaying active vaping (12.23 seconds) was significantly shorter ($P<.001$) than that of videos without active vaping (30.60 seconds). However, videos displaying nicotine sickness symptoms were longer in duration (28.57 seconds) than those not displaying nicotine sickness symptoms (19.92 seconds), albeit the comparison did not meet the threshold set for statistical significance ($P=.08$).

Users who posted first- or secondhand experiences with nicotine poisoning often did not explicitly or implicitly discourage others from engaging in harmful vaping behaviors such as overconsumption. Instead, users posting with the hashtag #nicsick appeared to actively participate in trending vape tricks or challenges. Some of the phrases heard in TikTok videos included “nic-sick check” and “where’s my JUUL” while users intentionally engaged in vaping behavior to generate TikTok trends (ie, how TikTok uses a hashtag or a song to group together viral trends within the app).

Discussion

Principal Findings

We found that TikTok users openly discussed their experiences, both firsthand and secondhand, with nicotine poisoning and related adverse events via the #nicsick hashtag. This includes users promoting extreme vaping behavior, discussion of users’ direct experiences with adverse events, and explicit video documentation of adverse events that appear to elicit user attention and reaction. We also found that users were more likely to create original videos as opposed to stitches and duets, which allow users to imbed and combine existing videos into their own TikTok post, which can often result in higher user engagement [34]. User engagement metrics of the videos in our study indicated moderate and lower compared to general metrics for popular TikTok videos, although some popular creators have had higher levels of engagement and mean follower counts. Additionally, the duration of videos depicting adverse events or nicotine poisoning symptoms was marginally longer than those that did not, indicating that videos depicting adverse events may need to be longer in order to provide sufficient details on the reported nicotine sickness experiences versus videos that simply reported or portrayed general vaping behavior.

These emerging forms of microvideo-based social interactions have the potential to change ways in which susceptible youth and young adults receive vaping-related behavior cues. For example, phrases heard in TikTok videos in this study included “nic-sick check” and “where’s my JUUL” and were associated with intentional overconsumption of nicotine products that also appeared to increase user engagement. TikTok’s *For You Page* also personalizes content based on user viewing history and can lead to a continuous cycle of exposure to trending vaping videos, including those in this study that portrayed overconsumption. Characteristics of TikTok videos, including use of hashtags, trending sounds, and video duration, may also represent crucial opportunities for the introduction of targeted health education and promotion to warn young users about the documented harms of vaping, particularly if they are already experiencing and self-reporting adverse events [35]

Yet, it is important to note that the impact of these videos on actual knowledge, attitudes, and behaviors of social media users associated with the appeal, uptake, and use of electronic nicotine delivery systems remains unclear. Vaping themes detected in this study seem to demonstrate that users may be promoting harmful vaping practices, which in turn has the potential to encourage overconsumption or dangerous nicotine use behavior. For example, popular internet challenges, including vape challenges, are rapidly circulated through TikTok—some may greatly harm the health of participants or encourage others to engage in similar behavior [25]. However, the motivations of these users to post #nicsick videos and the reason why they may or may not become viral require further study, particularly in the context of how the risk of electronic nicotine delivery system–related adverse events may be increased. We also found that there was no clear indication that users experiencing nicotine sickness were committed to quitting despite their sickness or poisoning, although a few user videos included individuals expressing a desire to quit. Instead, users using #nicsick appeared to be actively participating in trending vape tricks or challenges, which then led to overconsumption followed by one or more adverse events, which they then posted.

Combating the youth vaping epidemic requires scrutiny of how new interactive media, such as TikTok, serves as a source of vaping-related information and a source of peer influence among young user communities. Traditional poisoning and adverse event surveillance systems, such as the National Poison Data System and the US National Electronic Injury Surveillance System, may fail to capture these less severe cases captured on social media platforms, thereby underrepresenting the morbidity burden of nicotine poisoning [19]. Hence, social media platforms such as TikTok, may represent an important data source that can be used to help identify the burden of user-reported nicotine exposure and poisoning.

Our study narrowly focused on identifying and characterizing how nicotine poisoning–related content is created and shared through TikTok’s microvideo posts. Future studies should focus on examining the complex interplay between user exposure and the appeal of TikTok pro–tobacco and vaping content, the dynamics of how this content is shared and propagated on these networks, how it impacts real-world use patterns and vaping behavior, and finally, how these factors may impact safety and

health issues, such as adverse events. Though TikTok's policies state that content which offers the purchase, sale, trade, or solicitation of tobacco products (including vaping products) is prohibited [36]. Our study results indicate that content related to excessive or potentially dangerous smoking or vaping behavior does not fit this criterion, although the utility of removing such content is also unclear.

Limitations

We used a single platform and a fixed data collection period. Hence, the results may not be representative of all nicotine sickness content on social media. While TikTok is a popular social media platform, the demographics of TikTok users may not reflect those of the general population of electronic nicotine delivery system users. Furthermore, study data were limited to results for videos queried and returned by TikTok's internal search function, which may further limit generalizability. We used a single hashtag, #nicsick, based on a combination of open-source news reports, manual searches, and a preliminary Google Trends analysis to justify inclusion and due to the narrow focus of the study on nicotine sickness-related content. Hence, this study might have missed capturing additional nicotine sickness content with other hashtags that may have focused on specific electronic nicotine delivery system products, vaping behavior and challenges, and other nicotine and vaping-related topics with content related to nicotine sickness or adverse events. Furthermore, we did not cross-validate user-generated content with other data sources of nicotine sickness reports or cases, such as data from poison control

centers, and we did not explore user interaction or social network characteristics of web-based user networks where videos with nicotine poisoning content was shared, viewed, or interacted with, via favorites or comments, in-depth. Future studies should conduct more comprehensive and large-scale data analysis of nicotine-related adverse events on TikTok compared with those on other social media platforms. Future research can explore further analysis of characteristics of videos including hashtags, trending sounds, and video duration, which may help curate targeted health education and promotion against vaping use.

Conclusion

While we observed discussions and experiences with nicotine poisoning and adverse events on TikTok, further research is needed to assess how this unique social media risk environment impacts user perceptions about the known harms of vaping and which health promotion strategies can be tailored for youth experiencing vaping-related adverse events. Although other electronic nicotine delivery system-promoting content on TikTok may influence tobacco use behavior, #nicsick-specific content represents user-generated experiences directly associated with adverse events or poisoning due to nicotine overconsumption, which could be an alternative form of surveillance to better characterize this likely underreported public health issue. Supplementing traditional surveillance methods with infodemiology and infoveillance approaches can help in capturing the nicotine sickness cases that are not reported to poison control centers and assess the growing health burden due to nicotine use more appropriately.

Acknowledgments

This study was funded by the University of California Tobacco-related Disease Research Program (T29IP0465 and T29IP0384). We thank Cortni Bardier for her useful discussion, input, and assistance during the project.

Data Availability

Raw data supporting the conclusions of this paper will be made available by the authors, without undue reservation.

Authors' Contributions

ZL collected data. VP, TM, and MN conducted data analyses. All authors contributed to the design, formulation, drafting, completion, and approval of the final manuscript.

Conflicts of Interest

TKM serves as editor in chief for JMIR Infodemiology. ZL, MN, and TKM are employees of the startup company S-3 Research LLC. S-3 Research is a startup funded and currently supported by the National Institutes of Health National Institute of Drug Abuse through a Small Business Innovation and Research contract for opioid-related social media research and technology commercialization.

References

1. TikTok revenue and usage statistics. Business of Apps. URL: <https://www.businessofapps.com/data/tik-tok-statistics/> [accessed 2022-03-17]
2. Comp G, Dyer S, Gottlieb M. Is TikTok the next social media frontier for medicine? AEM Educ Train 2021 Jul 21;5(3):1-1 [FREE Full text] [doi: [10.1002/aet2.10532](https://doi.org/10.1002/aet2.10532)] [Medline: [34095694](https://pubmed.ncbi.nlm.nih.gov/34095694/)]
3. Omar B, Dequan W. Watch, share or create: the influence of personality traits and user motivation on TikTok mobile video usage. Int J Interact Mob Technol 2020 Mar 12;14(04):121. [doi: [10.3991/ijim.v14i04.12429](https://doi.org/10.3991/ijim.v14i04.12429)]
4. Chen MKY, Garden F, Sebaratnam DF. Isotretinoin on TikTok™: misinformation getting under our skin. Clin Exp Dermatol 2021 Dec;46(8):1606-1607. [doi: [10.1111/ced.14810](https://doi.org/10.1111/ced.14810)] [Medline: [34155690](https://pubmed.ncbi.nlm.nih.gov/34155690/)]

5. Villanti AC, Johnson AL, Ilakkuvan V, Jacobs MA, Graham AL, Rath JM. Social media use and access to digital technology in US young adults in 2016. *J Med Internet Res* 2017 Jun 07;19(6):e196 [FREE Full text] [doi: [10.2196/jmir.7303](https://doi.org/10.2196/jmir.7303)] [Medline: [28592394](https://pubmed.ncbi.nlm.nih.gov/28592394/)]
6. Montag C, Yang H, Elhai JD. On the psychology of TikTok use: a first glimpse from empirical findings. *Front Public Health* 2021;9:641673 [FREE Full text] [doi: [10.3389/fpubh.2021.641673](https://doi.org/10.3389/fpubh.2021.641673)] [Medline: [33816425](https://pubmed.ncbi.nlm.nih.gov/33816425/)]
7. Sun T, Lim C, Chung J, Cheng B, Davidson L, Tisdale C, et al. Vaping on TikTok: a systematic thematic analysis. *Tob Control* 2021 Jul 26. [doi: [10.1136/tobaccocontrol-2021-056619](https://doi.org/10.1136/tobaccocontrol-2021-056619)] [Medline: [34312317](https://pubmed.ncbi.nlm.nih.gov/34312317/)]
8. Roumeliotis G, Yang Y, Wang E, Alper A. Exclusive: U.S. opens national security investigation into TikTok - sources. Reuters. URL: <https://www.reuters.com/article/us-tiktok-cfius-exclusive/exclusive-u-s-opens-national-security-investigation-into-tiktok-sources-idUSKBN1XB4IL> [accessed 2022-01-29]
9. Auxier B, Anderson M. A majority of Americans say they use YouTube and Facebook, while use of Instagram, Snapchat and TikTok is especially common among adults under 30. Pew Research Center. 2021 Apr 07. URL: <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/> [accessed 2022-01-29]
10. Dai H, Hao J. Online popularity of JUUL and Puff Bars in the USA: 2019-2020. *Tob Control* 2022 Jan 13;31(1):7-10. [doi: [10.1136/tobaccocontrol-2020-055727](https://doi.org/10.1136/tobaccocontrol-2020-055727)] [Medline: [33051277](https://pubmed.ncbi.nlm.nih.gov/33051277/)]
11. King BA, Jones CM, Baldwin GT, Briss PA. The EVALI and youth vaping epidemics — implications for public health. *N Engl J Med* 2020 Feb 20;382(8):689-691. [doi: [10.1056/nejmp1916171](https://doi.org/10.1056/nejmp1916171)]
12. Chadi N, Hadland S, Harris S. Understanding the implications of the "vaping epidemic" among adolescents and young adults: a call for action. *Subst Abus* 2019;40(1):7-10 [FREE Full text] [doi: [10.1080/08897077.2019.1580241](https://doi.org/10.1080/08897077.2019.1580241)] [Medline: [30883295](https://pubmed.ncbi.nlm.nih.gov/30883295/)]
13. Hruska J, Maresova P. Use of social media platforms among Adults in the United States—behavior on social media. *Societies* 2020 Mar 23;10(1):27. [doi: [10.3390/soc10010027](https://doi.org/10.3390/soc10010027)]
14. Perez S. TikTok is being used by vape sellers marketing to teens. TechCrunch. 2021 Jan 26. URL: <https://techcrunch.com/2021/01/26/tiktok-is-being-used-by-vape-sellers-marketing-to-teens/> [accessed 2022-01-31]
15. Outbreak of lung injury associated with the use of e-cigarette, or vaping, products. Center for Disease Control and Prevention. 2020 Feb 25. URL: https://www.cdc.gov/tobacco/basic_information/e-cigarettes/severe-lung-disease.html [accessed 2022-01-31]
16. Nicotine poisoning. Cleveland Clinic. URL: <https://my.clevelandclinic.org/health/diseases/21582-nicotine-poisoning> [accessed 2022-01-29]
17. Abafalvi L, Péntzes M, Urbán R, Foley KL, Kaán R, Kispélyi B, et al. Perceived health effects of vaping among Hungarian adult e-cigarette-only and dual users: a cross-sectional internet survey. *BMC Public Health* 2019 Mar 13;19(1):302 [FREE Full text] [doi: [10.1186/s12889-019-6629-0](https://doi.org/10.1186/s12889-019-6629-0)] [Medline: [30866897](https://pubmed.ncbi.nlm.nih.gov/30866897/)]
18. Miller A. Nicotine poisoning increase due to e-cigarettes. *Can Med Assoc J* 2014 Jul 08;186(10):E367-E367 [FREE Full text] [doi: [10.1503/cmaj.109-4818](https://doi.org/10.1503/cmaj.109-4818)] [Medline: [24914118](https://pubmed.ncbi.nlm.nih.gov/24914118/)]
19. Chang JT, Wang B, Chang CM, Ambrose BK. National estimates of poisoning events related to liquid nicotine in young children treated in US hospital emergency departments, 2013-2017. *Inj Epidemiol* 2019 Apr 1;6(1):10 [FREE Full text] [doi: [10.1186/s40621-019-0188-9](https://doi.org/10.1186/s40621-019-0188-9)] [Medline: [31245259](https://pubmed.ncbi.nlm.nih.gov/31245259/)]
20. Xie J, Liu X, Dajun Zeng D. Mining e-cigarette adverse events in social media using Bi-LSTM recurrent neural network with word embedding representation. *J Am Med Inform Assoc* 2018 Jan 01;25(1):72-80 [FREE Full text] [doi: [10.1093/jamia/ocx045](https://doi.org/10.1093/jamia/ocx045)] [Medline: [28505280](https://pubmed.ncbi.nlm.nih.gov/28505280/)]
21. What it means to be "nic-sick". American Lung Association. 2019 Oct 01. URL: <https://www.lung.org/blog/nic-sick> [accessed 2022-01-29]
22. Basch C, Fera J, Pierce I, Basch C. Promoting mask use on TikTok: descriptive, cross-sectional study. *JMIR Public Health Surveill* 2021 Feb 12;7(2):e26392 [FREE Full text] [doi: [10.2196/26392](https://doi.org/10.2196/26392)] [Medline: [33523823](https://pubmed.ncbi.nlm.nih.gov/33523823/)]
23. Ostrovsky AM, Chen JR. TikTok and its role in COVID-19 information propagation. *J Adolesc Health* 2020 Nov;67(5):730 [FREE Full text] [doi: [10.1016/j.jadohealth.2020.07.039](https://doi.org/10.1016/j.jadohealth.2020.07.039)] [Medline: [32873499](https://pubmed.ncbi.nlm.nih.gov/32873499/)]
24. Baumel NM, Spatharakis JK, Karitsiotis ST, Sellas EI. Dissemination of mask effectiveness misinformation using TikTok as a medium. *J Adolesc Health* 2021 May;68(5):1021-1022. [doi: [10.1016/j.jadohealth.2021.01.029](https://doi.org/10.1016/j.jadohealth.2021.01.029)] [Medline: [33814279](https://pubmed.ncbi.nlm.nih.gov/33814279/)]
25. Kriegel E, Lazarevic B, Athanasian C, Milanaik R. TikTok, Tide pods and Tiger King: health implications of trends taking over pediatric populations. *Curr Opin Pediatr* 2021 Feb 01;33(1):170-177. [doi: [10.1097/MOP.0000000000000989](https://doi.org/10.1097/MOP.0000000000000989)] [Medline: [33337608](https://pubmed.ncbi.nlm.nih.gov/33337608/)]
26. Miller B, Meshnick A, Lushniak B. Electronic nicotine delivery systems: the need for continued regulatory innovation. *Tob Control* 2021 Aug 13:1. [doi: [10.1136/tobaccocontrol-2021-056622](https://doi.org/10.1136/tobaccocontrol-2021-056622)] [Medline: [34389688](https://pubmed.ncbi.nlm.nih.gov/34389688/)]
27. Kwon M, Park E. Perceptions and sentiments about electronic cigarettes on social media platforms: systematic review. *JMIR Public Health Surveill* 2020 Jan 15;6(1):e13673 [FREE Full text] [doi: [10.2196/13673](https://doi.org/10.2196/13673)] [Medline: [31939747](https://pubmed.ncbi.nlm.nih.gov/31939747/)]
28. Kong W, Song S, Zhao Y, Zhu Q, Sha L. TikTok as a health information source: assessment of the quality of information in diabetes-related videos. *J Med Internet Res* 2021 Sep 01;23(9):e30409 [FREE Full text] [doi: [10.2196/30409](https://doi.org/10.2196/30409)] [Medline: [34468327](https://pubmed.ncbi.nlm.nih.gov/34468327/)]

29. Song S, Zhao Y, Yao X, Ba Z, Zhu Q. Short video apps as a health information source: an investigation of affordances, user experience and users' intention to continue the use of TikTok. *Internet Res* 2021 Jun 24;31(6):2120-2142. [doi: [10.1108/intr-10-2020-0593](https://doi.org/10.1108/intr-10-2020-0593)]
30. Tan ASL, Weinreich E. #PuffBar: how do top videos on TikTok portray Puff Bars? *Tob Control* 2021 Nov;30(6):712-713. [doi: [10.1136/tobaccocontrol-2020-055970](https://doi.org/10.1136/tobaccocontrol-2020-055970)] [Medline: [32934090](https://pubmed.ncbi.nlm.nih.gov/32934090/)]
31. Wang B, Liu S, Persoskie A. Poisoning exposure cases involving e-cigarettes and e-liquid in the United States, 2010-2018. *Clin Toxicol (Phila)* 2020 Jun 2;58(6):488-494 [FREE Full text] [doi: [10.1080/15563650.2019.1661426](https://doi.org/10.1080/15563650.2019.1661426)] [Medline: [31496321](https://pubmed.ncbi.nlm.nih.gov/31496321/)]
32. Rutherford BN, Sun T, Johnson B, Co S, Lim TL, Lim CCW, et al. Getting high for likes: exploring cannabis-related content on TikTok. *Drug Alcohol Rev* 2022 Jan 24;1. [doi: [10.1111/dar.13433](https://doi.org/10.1111/dar.13433)] [Medline: [35073422](https://pubmed.ncbi.nlm.nih.gov/35073422/)]
33. Carter PN, Hall EE, Ketcham CJ, Ahmed OH. Not just for dancing? a content analysis of concussion and head injury videos on TikTok. *Front Sports Act Living* 2021 Oct 27;3:692613 [FREE Full text] [doi: [10.3389/fspor.2021.692613](https://doi.org/10.3389/fspor.2021.692613)] [Medline: [34778757](https://pubmed.ncbi.nlm.nih.gov/34778757/)]
34. MacKinnon KR, Kia H, Lacombe-Duncan A. Examining TikTok's potential for community-engaged digital knowledge mobilization with equity-seeking groups. *J Med Internet Res* 2021 Dec 09;23(12):e30315 [FREE Full text] [doi: [10.2196/30315](https://doi.org/10.2196/30315)] [Medline: [34889739](https://pubmed.ncbi.nlm.nih.gov/34889739/)]
35. Fedt A, Bhattarai S, Oelstrom M. Vaping-associated lung injury: a new cause of acute respiratory failure. *J Adolesc Health* 2020 Jun;66(6):754-757. [doi: [10.1016/j.jadohealth.2020.03.019](https://doi.org/10.1016/j.jadohealth.2020.03.019)] [Medline: [32473722](https://pubmed.ncbi.nlm.nih.gov/32473722/)]
36. Community guidelines. TikTok. 2022 Feb. URL: <https://www.tiktok.com/community-guidelines?lang=en> [accessed 2022-01-29]

Edited by A Mavragani; submitted 04.10.21; peer-reviewed by S Song, JP Allem, H Zhao; comments to author 08.12.21; revised version received 01.02.22; accepted 04.02.22; published 30.03.22.

Please cite as:

Purushothaman V, McMann T, Nali M, Li Z, Cuomo R, Mackey TK

Content Analysis of Nicotine Poisoning (Nic Sick) Videos on TikTok: Retrospective Observational Infodemiology Study
J Med Internet Res 2022;24(3):e34050

URL: <https://www.jmir.org/2022/3/e34050>

doi: [10.2196/34050](https://doi.org/10.2196/34050)

PMID: [35353056](https://pubmed.ncbi.nlm.nih.gov/35353056/)

©Vidya Purushothaman, Tiana McMann, Matthew Nali, Zhuoran Li, Raphael Cuomo, Tim K Mackey. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 30.03.2022. 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.

Original Paper

Exploring the Discursive Emphasis on Patients and Coaches Who Participated in Technology-Assisted Diabetes Self-management Education: Clinical Implementation Study of Health360x

Muhammed Y Idris¹, PhD; Ernest Alema-Mensah², BSc, MSc, DMin, PhD; Elizabeth Olorundare³, MPH, MD; Mohammad Mohammad³, MD; Michelle Brown⁴, BSc, MPH; Elizabeth Ofili⁵, MPH, MD; Priscilla Pemu^{3,5}, MSc, MD

¹Department of Environmental Medicine and Public Health, Icahn School of Medicine at Mount Sinai, New York, NY, United States

²Department of Community Health & Preventive Medicine, Morehouse School of Medicine, Atlanta, GA, United States

³Clinical Research Center, Morehouse School of Medicine, Atlanta, GA, United States

⁴Morehouse Choice Accountable Care Organization and Education System, Inc, Morehouse School of Medicine, Atlanta, GA, United States

⁵Department of Medicine, Morehouse School of Medicine, Atlanta, GA, United States

Corresponding Author:

Priscilla Pemu, MSc, MD

Department of Medicine and Clinical Research Center

Morehouse School of Medicine

720 Westview Drive SW

Atlanta, GA, 30310

United States

Phone: 1 4047566684

Fax: 1 4047561328

Email: pipemu@msm.edu

Abstract

Background: A critical unmet need for underserved patients with diabetes is regular access to sufficient support for diabetes self-management. Although advances in digital technologies have made way for eHealth applications that provide a scalable path for tailored interventions for self-management of chronic conditions, health and digital literacy has remained an obstacle to leveraging these technologies for effective diabetes self-management education. Studies have shown that the availability of coaches helps to maintain engagement in internet-based studies and improves self-efficacy for behavior change. However, little is known about the substances involved in these interactions.

Objective: This study aims to compare the content of conversations between patient-coach pairs that achieved their self-management goals and those that did not. The context is a clinical implementation study of diabetes self-management behavior change using Health360x within the practices of the Morehouse Choice Accountable Care Organization in the Atlanta metro area. Health360x is a coach-assisted consumer health information technology designed to support self-management skills acquisition and behavior among underserved, high-risk patients with diabetes.

Methods: We provide a novel analysis of the discursive emphasis on patients and coaches. We examined transcripts of visits using a structural topic model to estimate topic content and prevalence as a function of patient and coach characteristics. We compared topics between patient-coach pairs that achieved diabetes-related self-management goals and those who did not. We also estimated a regression in which utterances are the units, the dependent variable is the proportion of an utterance that is about a given topic, and the independent variables are speaker types and explored other themes.

Results: Transcripts from 50 patients who were recruited and consented, starting in February 2015, were analyzed. A total of 44 topics were estimated for patient-coach pairs that achieved their intended health goals and 50 topics for those who did not. Analysis of the structural topic model results indicated that coaches in patient-coach pairs that were able to achieve self-management goals provided more contextual feedback and probed into patients' experience with technology and trust in consumer information technologies. We also found that discussions around problem areas and stress, support ($\beta_{\text{Coach}}=.015$; $P<.001$), initial visits ($\beta_{\text{Coach}}=.02$; $P<.001$), problems with technology ($\beta_{\text{Coach}}=.01$; $P<.001$), health eating goals ($\beta_{\text{Coach}}=.01$; $P=.04$), diabetes knowledge

($\beta_{\text{Coach}}=.02$; $P<.001$), managing blood sugar ($\beta_{\text{Coach}}=.03$; $P<.001$), and using Health360x ($\beta_{\text{Coach}}=.003$; $P=.03$) were dominated by coaches.

Conclusions: Coach-facilitated, technology-based diabetes self-management education can help underserved patients with diabetes. Our use of topic modeling in this application sheds light on the actual dynamics in conversations between patients and coaches. Knowledge of the key elements for successful coach–patient interactions based on the analysis of transcripts could be applied to understanding everyday patient–provider encounters, given the recent paradigm shift around the use of telehealth.

(*J Med Internet Res* 2022;24(3):e23535) doi:[10.2196/23535](https://doi.org/10.2196/23535)

KEYWORDS

self-management; structural topic modeling; coaching; diabetes; minority populations; mobile phone

Introduction

e-Patients are a relatively new breed of informed health consumers comprising 61% of all adults in the United States using technology to *equip, enable, empower, and engage* with relevant health-related information [1]. Despite increasing access to wireless and mobile technology among minority and vulnerable populations, there is still a higher trend in the adoption of digital health technologies among disproportionately advantaged groups, such as young, White males [1-3]. It is known and well-accepted in the disparities research community that health and digital literacy are barriers to the increased use of eHealth technologies among minority and vulnerable populations. These obstacles include lack of perceived value, technologies creating more work, materials not in appropriate reading levels to make informed health decisions, lack of cultural relevance (ie, accessing sources in their respective language), and privacy and trust concerns [4]. These barriers are a function of individual characteristics, including socioeconomic status, education, and less access to change agents, among other factors that are associated with higher rates of attrition in internet-based studies [5-7]. Higher rates of attrition are even more noticeable among minority and underserved populations, who are also disproportionately challenged by intervention factors, including usability problems, complexity, and ease of discontinuance [4]. The net result is a lack of diffusion of innovations among vulnerable populations and a worsening of disparities in health outcomes [8], that is, the uneven adoption of digital health innovations by minority and other vulnerable populations because of health and digital literacy exacerbates existing disparities in health.

An area where there is an agreement in the literature is that health and digital literacy are obstacles to the use of eHealth technologies for implementing tailored health interventions that support self-management of chronic conditions, particularly diabetes [7]. A typical example for diabetes is the challenge with the delivery of appropriate and effective diabetes self-management education (DSME) to vulnerable populations. Multiple studies have found that diabetes DSME is associated with improved diabetes knowledge and self-care behavior [9], improved clinical outcomes such as lower hemoglobin A_{1c} [10], lower self-reported weight [11,12], improved quality of life [13], healthy coping [13,14], and lower costs [15]. Diabetes education is associated with increased use of primary and preventive services and lower use of acute, inpatient hospital services [16], and patients who participate in DSME are more

likely to follow best practice treatment recommendations, particularly among the Medicare population. A critical issue and unmet need for underserved patients living with diabetes is the regular access to sufficient support for effective diabetes self-management. Better outcomes were reported for DSME interventions that were longer and included follow-up support, that were culturally appropriate and age appropriate, and that were tailored to the individual psychosocial needs and behavioral preferences [17,18].

Prior studies indicate that the availability of a trusted partner (coach) helps to improve user experience with health technology for DSME, including perceived relevance and ease of use [19], among vulnerable populations who face health and digital literacy barriers and higher rates of attrition. The support of a coach improves engagement with technology, and accountability improves adherence to and self-efficacy for behavior change [20]. In this patient–coach relationship, coaches are not intermediaries between patients and health information but serve as an apomediary, that is, a guide to relevant information and services for the patient [19]. Much has been theorized about what leads to successful patient–coach relationships; however, what is less understood is the content of patient–coach interactions.

In this paper, we aim to evaluate the discursive emphasis on patients and coaches engaged in DSME. To do so, we will leverage self-reported patient data and >100 hours of recorded transcripts from a clinical implementation study of Health360x, a coach-assisted consumer health information technology (CHIT) designed to support diabetes self-management skills acquisition and behavior for underserved, high-risk patients living with diabetes [19]. The platform is available as a web application and mobile app and includes functionality for improving health literacy and self-efficacy through access to a health coach, a social network of peers, a curriculum of DSME materials, and a health tracker that can record blood pressure, BMI, physical activity, and self-management goals.

In what follows, we provide some background on the study from which the data were generated. We then provide a novel analysis of the discursive emphasis on patients and coaches from transcripts of visits using a structural topic model (STM) to estimate topic content (ie, characteristic words in a topic) and prevalence (ie, proportion of an utterance that is related to a topic) as a function of patient and coach characteristics. We conclude with a discussion of our results, limitations, and

directions for future work on the use of coach-assisted CHIT for DSME.

Methods

Transdisciplinary Collaborative Center Project

Longitudinal data and transcripts of patient-coach interactions were taken from a clinical implementation study of Health360x, a coach-assisted CHIT designed to assist with chronic illness care through behavior change [19,21]. It is important to note that as an implementation study, there was no intervention or control arm as the goal of the study is not to make generalizations to a broader population [22]. The research design of the study focused almost exclusively on external validity. As such, countless hours were spent working with clinical practices through focus groups and plan-do-study-act cycles to ensure clinical readiness. The goal of the transdisciplinary collaborative center was to study the implementation across several clinics that are part of the Morehouse Choice Accountable Care Organization (MCACO).

Health360x was developed at the Morehouse School of Medicine and evaluated within the MCACO. MCACO is a physician-led integrated delivery model participating in the Medicare Shared Savings Program offered by the Centers for Medicare and Medicaid Services Innovation Center [23]. The MCACO partner practices care for a disproportionate share of high-need, complex populations and endure extraordinary challenges in managing the use, with comparatively limited resources. The Health360x implementation at MCACO was nested within the MCACO's Centralized Care Coordination model. This synergistic approach represents a *health system* where an African American majority patient population spends time and receives primary health care.

To determine health practice readiness for inclusion in the study, key individuals within the practice were identified for a focus group. For a clinic to participate, their front office manager, practice manager, lead physician, and (optionally) patients living with diabetes were required to participate in this focus group. The goal of this focus group was to delineate patient experiences and practice workflow for patients living with diabetes. Findings from focus group discussions were then used to identify and address issues around recruitment, signup, and visits through a plan-do-study-act cycle from select practices. In total, we recruited between 3 and 5 individuals across 5 sites for a total of 25 individuals for these focus groups.

Health coaches from within and outside of clinical practices were also recruited. The preferred characteristics included the following: (1) certified by Americans With Disabilities Act with the tool *Fundamentals of Diabetes Care*; (2) previous experience with living with diabetes either directly or through providing care for a family member or patient; (3) health care professionals with previous experience in educating patients living with diabetes; and (4) soft skills, including being considerate, responsible, dependable, and understanding within their community. A total of 4 health coaches were recruited by participating in health practices. This included 3 foreign-trained physicians, including 1 man in his late 30s whose second

language was English, and 2 female physicians in their late 30s who were native English speakers. The final coach was a licensed practical nurse in her late 50s. All coaches identified as Black, with 50% (2/4) of coaches identifying as African and the others as African American.

Health coaches interacted with a slate of patients identified through the MCACO practices and established communication with the practice to facilitate patient self-management support. Approximately 20-30 patients across 4 practices were selected, obtained consent from, and trained on the use of Health360x to report outcomes of self-efficacy; overall health, including blood pressure, blood glucose, exercise, sleep, and quality of life; and satisfaction with care. A total of 200 patients were screened for being at high risk of living with diabetes. A total of 100 patients were excluded as they did not meet the eligibility criteria for the study or were part of a clinical site that was ultimately deemed *not* practice ready, leaving 100 patients to be allocated the intervention.

Ethics Approval

This study was approved by the Morehouse School of Medicine Institutional Review Board (approval number 674).

Inclusion Criteria

Included in the study were adults living with diabetes at high risk of complications and mortality, as defined by the presence of obesity or overweight status ($BMI > 25 \text{ kg/m}^2$); tobacco use; history of depression; systolic blood pressure $> 140 \text{ mm Hg}$ and diastolic blood pressure $> 90 \text{ mm Hg}$; hemoglobin $A_{1c} > 7\%$; recent hospitalization or emergency room visit for uncontrolled diabetes or hypoglycemia; history of renal disease defined as estimated glomerular filtration rate $< 60 \text{ mg/mmol}$; and history of heart attack, angina, claudication, or cerebrovascular disease.

Training

Health coaches were required to complete 4 sessions of training. These sessions included training on diabetes knowledge, diabetes management and prevention, cultural competencies of the target population, working with low-literacy populations, patient recruitment and retention, confidentiality, and familiarity with data collection tools. Coaches were also required to complete web training and certification by the American Association of Diabetes Educators titled *Fundamentals of Diabetes Care*. Finally, the coach training sessions also included hands-on training on Health360x.

A total of 6 hours of training sessions were required to be completed by patients before the start of participation. The training was conducted at the Morehouse School of Medicine and covered the role of health coaches as navigators; the use of Health360x, glucose meter, blood pressure monitor, and pedometer; and process for downloading data. Patients with access to a computer or personal smartphone compatible with Health360x were instructed on how to access or download the app. Those patients who did not have access to a computer or personal smartphone that is compatible with the mobile app relied on a kiosk at the practice that was accessible when they visited their coach. The Health360x application access and training were completed through a combination of YouTube

videos, and a handbook detailing steps for using the application was also provided to patients.

Patient and Coach Visits

The health coach met with patients at the practice during scheduled visits. The purpose of each visit was to advance the patients' self-efficacy for self-management of their diabetes. During the initial visit, patients were asked to select individual diabetes-related self-management health goals around being active, healthy coping and eating, monitoring, problem solving, reducing risky behaviors, and taking medication. At each subsequent visit, coaches ascertained the barriers the patient experienced through the use of motivational interviewing. Once these barriers were identified, incremental steps to address them, which the patient suggested or agreed with, were made. Techniques such as behavioral contracting were incorporated into the process. All interactions between patients and coaches were between 20 and 40 minutes. All meetings were digitally recorded and transcribed using a third-party transcription service. As a rule, initial schedules for interactions were every 2 weeks for the first 2 months and then monthly for 4 months.

Statistical Analysis

Primary Outcome in Transdisciplinary Collaborative Center Study

The primary outcome of interest in the study was self-efficacy for behavior change, which is a predictor of intention and behavior. Self-efficacy was measured using a self-reported 10-point General Self-Efficacy Scale at each follow-up visit. A patient was determined to have achieved their self-efficacy goals if they self-reported a score ≥ 7 .

Sample Size

Our sample size calculation was based on a systematic review and meta-analysis that included 85 high-quality published studies of internet-based behavioral health interventions with a total sample size of 43,236 patients [24]. On average, interventions with more extensive use of theory were associated

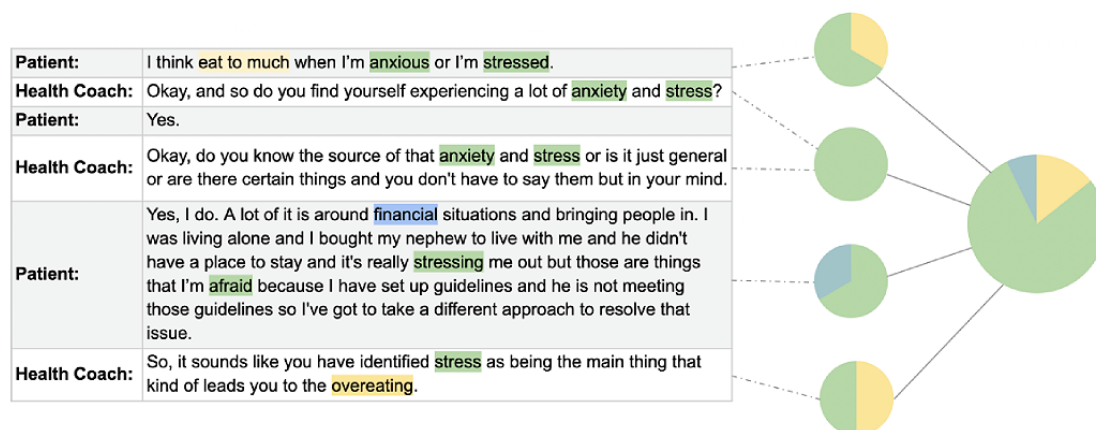
with increases in effect size ($P=.049$). Interventions based on the theory of planned behavior tended to have substantial effects on behavior (Cohen $d=0.36$, 95% CI 0.15-0.56). Health360x was designed with particular attention to the incorporation of theory-driven behavioral change techniques such as an enriched information environment (current diabetes knowledge and curriculum), self-monitoring with color-coded feedback, goal setting, identifying barriers, and problem solving. Our a priori power analysis indicated that a minimum sample size of 97 was required to detect an anticipated effect size of 0.36, assuming a desired power and significance level of 0.8 and 0.05, respectively. Therefore, we recruited 100 patients for this study.

Topic Modeling

To explore the substance of conversations between patients and coaches, we used a topic model. This statistical machine learning method identifies common themes in a corpus of documents, or in our case transcripts, and is based on the intuition that particular words are associated with particular topics. Formally, each utterance in a conversation is represented as a vocabulary multiset with corresponding word counts (also referred to as a document-term matrix). Each word in this vocabulary has some associated probability that it belongs to a given topic. If we think of a conversation as representing a random mixture of topics that are in turn defined by a set of characteristic, high-probability words, we can infer latent topics.

Topic modeling has been shown to reliably uncover conversational topics from transcripts between patients and providers in clinical settings [25,26]. Figure 1 provides a visual example of how this works from a snippet of the conversation between a patient and a health coach. Characteristic words are highlighted in yellow, green, and blue and correspond to eating, anxiety, and finances, respectively. Taken together, we can see that this conversation is about anxiety and stress, leading to overeating. However, the snippet can be further broken down, and a topic model estimates the proportion of the conversation that corresponds to each of these topics based on the observed vocabulary.

Figure 1. Example topic proportions and assignment.



We use an STM [27] for 2 main reasons: first, the STM extends traditional probabilistic models such as the Blei et al [28] latent

Dirichlet allocation and the Blei and Lafferty [29] Correlated Topic model to incorporate document metadata into the analysis

of latent topics. This was done by defining a generative process that produces *documents*, that is, utterances, with associated data, given document-topic and topic-word distributions, and using a variational optimization algorithm to estimate topic model parameters. Each conversation and utterance was made up of different proportions of the estimated topics. Incorporating speaker metadata into our topic modeling allowed us to explicitly model the association among patients, coaches, and topics of conversation, that is, the way the characteristics of patients and coaches were associated with the proportion of a specific topic in a given utterance. In our topic model, we included the speaker's age, gender, and race and ethnicity, as well as that of the person they are speaking with and whether the speaker was a coach. We estimated a regression in which utterances were the units, the dependent variable was the proportion of an utterance that is about a given topic, and the independent variables were speaker types.

The second reason we used the STM relates to what is a common subjective decision made when using topic models: the number of topics to estimate. This decision was generally arbitrary. To avoid any human-injected bias in our analysis, we leveraged a data-driven approach to select the number of topics based on the identification of *anchor* words that exist in a document only if the document is about a specific topic. The Mimno and Lee [30] algorithm automatically selects distinctive and probable anchor words from which a number of topics can be reliably estimated. As the authors of the STM indicate in their documentation of their software in R (R Foundation for

Statistical Computing), the primary advantage of this approach is that the number of topics to be estimated is automatically selected by the algorithm.

Results

Patient Sample Characteristics

A total of 100 high-risk patients with diabetes were recruited, and they consented to participate in this study from February 2015 over 5 years. Of the 100 patients, 18 (18%) patients were lost to follow-up before starting any coaching sessions, leaving 82 (82%) patients with at least 1 and as many as 11 follow-up visits. A total of 195 patient-coach follow-up visits were completed throughout the study, with an average of 2.35 (SD 1.98) follow-up visits per patient and 4.79 (SD 5.77) goals per patient. Of the 82 patients, 3 (4%) patients specified 1 diabetes-related self-management health goals, 18 (22%) patients specified 2 goals, 15 (18%) patients specified 3 goals, 8 (10%) patients specified 5 goals, 4 (5%) patients specified 7 goals, 1 (1%) patient specified 8 goals and another 15 goals, and 1 (1%) remaining patient specified 24 goals. Of the 82 patients who participated in the study, transcripts were analyzed for 51 (62%) patients. Digital recordings for the remaining 38% (31/82) of patients were of bad quality and, therefore, were not uploaded or corrupted during transcription. Most patients were Black and African American (45/51, 88%) and between the ages of 45 and 65 years (40/51, 78%). [Table 1](#) provides a complete breakdown of patient demographic characteristics. [Textbox 1](#) provides a sample of diabetes-related self-management goals.

Table 1. Patient demographic characteristics (N=51).

Characteristics	Values, n (%)
Age range (years)	
18-24	1 (2)
25-34	3 (6)
35-44	2 (4)
45-54	22 (43)
55-65	18 (35)
>65	3 (6)
No response	2 (4)
Gender	
Male	25 (49)
Female	26 (51)
Race and ethnic background	
Asian or Pacific Islander	1 (2)
Black or African American	45 (88)
Hispanic or Latino	1 (2)
White	2 (4)
Indian or Alaskan Native	1 (2)
No response	1 (2)
Insurance	
No insurance	17 (33)
Medicare	12 (24)
Medicaid	15 (29)
Dual eligible	5 (10)
No response	2 (4)

Textbox 1. Frequency of the number of health subgoals.

Health subgoals

- Exercise longer
- Cope with diabetes
- Get support from my medical team
- Get support from my family and friends
- Get preventive help
- Stop smoking
- Check my feet
- Lose weight
- Get blood pressure under control
- Learn to have a safe pregnancy
- Follow my eating schedule better
- Eat better food
- Overeat less often
- Check my blood sugar more often
- Miss fewer blood sugar checks
- Do my blood sugar checks on time more often
- Prevent high blood sugars
- Treat high blood sugars
- Prevent low blood sugars
- Treat low blood sugars
- Manage diabetes when sick
- Miss fewer medications

Structural Topic Model

Preprocessing

In total, our corpus included >17,000 talk turns, that is, an utterance or spoken statement in a conversation. To explore differences between topics of conversation between patient-coach pairs that achieved their health goals and those that did not, we split the corpus into successful patient-coach pairs (achieved) and patient-coach pairs that did not achieve the intended health goals (not achieved). This left us with 7196 talk turns in the achieved corpus and 9644 talk turns in the not achieved corpus. We then preprocessed each corpus using the `textProcessor` function in the STM R package to remove punctuation, stop words, numbers, and stemming. This produced a final document-term matrix with 6737 talk turns and a vocabulary of 1444 terms for the achieved group and 8663 talk turns and 1952 terms for the not achieved group. Patients represented more talk turns, that is, spoke more, in both the

achieved and the not achieved groups, 60% (4318/7196) and 63% (6076/9644), respectively.

Topical Content

A total of 44 topics were estimated for patient-coach pairs that achieved their intended health goals and 50 topics for those who did not. In line with common practice, our approach to assigning labels to topics began by reviewing topic model outputs to identify potential topics based on the coherence of characteristic word forms. We then reviewed the high-scoring conversations for each topic to determine the topic labels. Where there was a clear topical theme in the first N talk turns, we validated the topic label. In general, a clear theme emerged within the first 30 talk turns. A review of high-scoring talk turns was completed collaboratively by 4 reviewers until a consensus on topic labels was reached. Of the 44 topics estimated for the achieved group, 13 (29%) coherent topics emerged, as presented in [Textbox 2](#). [Textbox 3](#) provides the results for the second STM of the not achieved group, which produced 10 coherent topics.

Textbox 2. Structural topic model 1 for patient–coach pairs that achieved health goals.

Topic 3: healthy eating

- Eat
- Much
- Littl
- Meal
- Sometim
- Even
- Pretti
- Plan
- Ive
- Cook

Topic 5: monitoring blood pressure and blood sugar

- One
- Ten
- Twenti
- Hundr
- Fifteen
- Forti
- Twentieth
- Salad

Topic 6: healthy coping

- Diabet
- Feel
- Make
- Tell
- Confid
- Can
- Manag
- Enough
- Clear
- Prepar

Topic 13: use of consumer health information technology

- Use
- Password
- Ask
- Year
- Answer
- Comput
- Yesterday
- Long
- Past

Topic 18: healthy snacking

- Like
- Look
- Healthi
- Snack
- Always
- Challeng
- Per
- Made
- God
- Seem

Topic 22: goal setting and habits

- Want
- Mean
- Lot
- Youv
- Hous
- Line
- Note
- Never
- That'

Topic 25: Health Insurance Portability and Accountability Act and diabetes knowledge (initial visit)

- Health
- Inform
- Twelv
- Follow
- Account
- Reduc
- Accept
- Moham
- Higher
- Behavior

Topic 26: sleep

- Well
- Start
- Sleep
- Did not
- Went
- Bed
- Oclock
- Fill
- Sound

- Hurt

Topic 30: monitoring for diabetics

- Blood
- Sugar
- Check
- Morn
- Test
- High
- Prevent
- Weight
- Urin
- Cholesterol

Topic 31: follow-up visit

- Two
- Come
- Week
- Back
- Month
- That's
- Appoint
- Thursday
- Within
- You will

Topic 38: diabetes knowledge test

- Glucos
- Exercis
- Insulin
- Low
- Blood
- Juic
- Lower
- Drink
- Fat
- Effect

Topic 43: health coping and strategies (coaching)

- Problem
- Stress
- Situat
- Serious
- Barrier
- Somewhat
- Live

- Import
- Way
- Facebook

Topic 44: goal assessment and progress (coaching)

- Said
- Anyth
- Continu
- Modifi
- Were not
- Alright
- Tomorrow
- Say
- That'
- Ive

Textbox 3. Structural topic model for patient–coach pairs that did not achieve health goals.

Topic 1: medication

- Take
- Medic
- Medicin
- Side
- Took
- Wednesday
- Taken
- Have not
- Does not
- Gave

Topic 2: problem areas, and stress and support

- Feel
- Diabet
- Confid
- Umm
- Hot
- Track
- Level
- Control
- Comfort
- Live

Topic 9: Health Insurance Portability and Accountability Act and diabetes knowledge

- Help
- Health
- Care
- Inform
- Plan
- Idea
- Access
- Program
- Save
- Studi

Topic 14: problems with technology

- Use
- Way
- Password
- Pretti
- Email
- Step
- Happen
- Ahead

- Open
- Somewher

Topic 15: healthy eating goals

- Eat
- Hmm
- Goal
- Food
- Bettwe
- Set
- Schedul
- Term
- Active
- Healthi

Topic 20: diabetes knowledge test

- Blood
- Much
- Glucos
- Pressur
- Insulin
- Test
- Drink
- Enough
- Juic
- Check

Topic 21: social determinants

- Year
- Type
- Famili
- Ago
- Friend
- Person
- Meet
- Hes
- Son
- Brother

Topic 22: sleep

- Might
- Sleep
- Night
- Bed
- Monday
- Past

- Friday
- Somebody
- Wake
- Troubl

Topic 44: managing blood sugar (coaching)

- Sugar
- Also
- High
- Blood
- Low
- Read
- Class
- Prevent
- Knowlegd
- Impact

Topic 49: using Health360x

- Can
- Put
- Now
- Enter
- Sheet
- Consent
- Link
- Pro
- Mix
- Anytim

Coaching and Contextual Feedback

We found quite a bit of overlap between topics of discussion between the 2 groups (eg, healthy eating, diabetes knowledge, and sleep). However, we found more coherent, that is, meaningful, topics that capture coaching in the achieved group versus in the not achieved group. Further investigation of high-scoring talk turns and the broader conversation that they were a part of provides many examples of coaches providing contextual feedback and incorporating this feedback into the patient's strategy. [Multimedia Appendix 1](#) provides 2 examples of such conversations. In the first example, which is taken from a high-scoring conversation on healthy coping and strategies, we see that after the coach first probes the source of anxiety and stress that is leading to overeating, acknowledges the patient's feelings, and contextualizes how their anxiety and stress lead to overeating, only then does the coach start discussing a plan to reduce overeating by focusing on the identified stressors. The next sample conversation was around healthy eating. The coach begins the conversation by recentering the patient around the goal of getting their blood sugar down

through exercise and healthier eating habits. The conversation quickly zeroes in a specific challenge in implementing this strategy: not eating on a regular schedule as the patient was not hungry at regular mealtimes. Only then does the coach provide contextualized feedback around what, when, and how much to eat.

User Experiences With CHIT

Another clear theme in our analysis of the topic model results was conversations around user experiences with CHIT. We found that CHIT use in the not achieved group focused almost exclusively on technical issues and how to use Health360x. [Multimedia Appendix 2](#) provides 2 examples from the not achieved group. In the first example, the coach expresses frustration with technical issues that require troubleshooting Health360x during a demonstration with a patient. The second example is from a conversation in which the patient and coach work together to reset the password (a common theme throughout the study was remembering passwords) and then updating the patient's health record on Health360x.

Conversely, the topic on the use of CHIT in the achieved group contained multiple examples of coaches probing user experiences with technology. [Multimedia Appendix 3](#) provides 3 examples of coaches asking probing questions on CHIT use in the achieved group. These questions ranged from familiarity with technology and use of the internet to how trusting the patients are of consumer information technologies, including sharing their health and other information on those tools. What we found to be particularly interesting is that we did not find a topic in the achieved group that resembled issues with using Health360. This suggests that understanding a patient's experiences with technology can lead to better coaching around the use of technology. This, in turn, can lead to better engagement with CHIT, self-management behavior changes, and health outcomes.

Who Dominates Topics of Conversation: Patients or Coaches?

In the achieved group, we found that healthy eating was much more likely to be discussed by patients. The average proportion of conversations about healthy eating across all patient and coach interactions, that is, E (*topic prevalence*), was 0.026 or 2.6%. Being a coach was associated with an increase of 4.9% ($\beta_{\text{Coach}}=.049$; $P=.02$) in the healthy eating topic proportion of a given utterance. The average proportion of conversations around follow-up visits and coaching in the form of goal assessment and progress across all visits was 3.4% and 0.8%, respectively. Being a coach was associated with an increase of 3% ($\beta_{\text{Coach}}=.03$; $P=.007$) and 0.3% ($\beta_{\text{Coach}}=.003$; $P=.01$) in the proportion of utterances related to follow-up visits and coaching in the form of goal assessment and progress. All other coherent topics in the achieved group were equally as likely to be discussed by either coaches or patients.

Conversations in the not achieved group were dominated by coaches. This included a 1.5% increase in the proportion of an utterance on problem areas, and stress and support ($\beta_{\text{Coach}}=.015$; $P<.001$; E [*topic prevalence*]=0.012), a 2.4% increase related to initial visits ($\beta_{\text{Coach}}=.024$; $P<.001$; E [*topic prevalence*]=0.012), a 1.3% increase around problems with technology ($\beta_{\text{Coach}}=.013$; $P<.001$; E [*topic prevalence*]=0.012), a 0.8% increase in health eating goals ($\beta_{\text{Coach}}=.008$; $P=.04$; E [*topic prevalence*]=0.03), a 1.7% increase in the proportion of diabetes knowledge ($\beta_{\text{Coach}}=.017$; $P<.001$; E [*topic prevalence*]=0.02), a 3% increase in speaking about managing blood sugar ($\beta_{\text{Coach}}=.03$; $P<.001$; E [*topic prevalence*]=0.02), and a 0.3% increase in speaking about using Health360x ($\beta_{\text{Coach}}=.003$; $P=.03$; E [*topic prevalence*]=0.03). A total of 2 exceptions were the proportion of an utterance on medication ($\beta_{\text{Coach}}=-0.008$; $P=.009$; E [*topic prevalence*]=0.02) and social determinants ($\beta_{\text{Coach}}=-0.02$; $P<.001$; E [*topic prevalence*]=0.012), which were positively associated with patients.

[Figures 2](#) and [3](#) present the mean difference in topic proportions as a function of whether the speaker was a coach or a patient, that is, which topics were more likely to be discussed by coaches versus patients. Also included in the plots are CIs indicating a statistically significant difference between the 2 groups. Although we found quite a bit of overlap between topics of discussion between the 2 groups, plotting the mean difference in topic proportions across all coherent topics indicated that conversations in the achieved group were equally distributed between patients and coaches. However, topics of conversation between patients and coaches in the not achieved group did not include 0 in their 95% CIs, indicating that these topics were dominated, and therefore more likely to be driven, by coaches.

Figure 2. Mean difference in achieved topic proportions—patients versus coaches. BP: blood pressure; HIPAA: Health Insurance Portability and Accountability Act; IT: information technology.

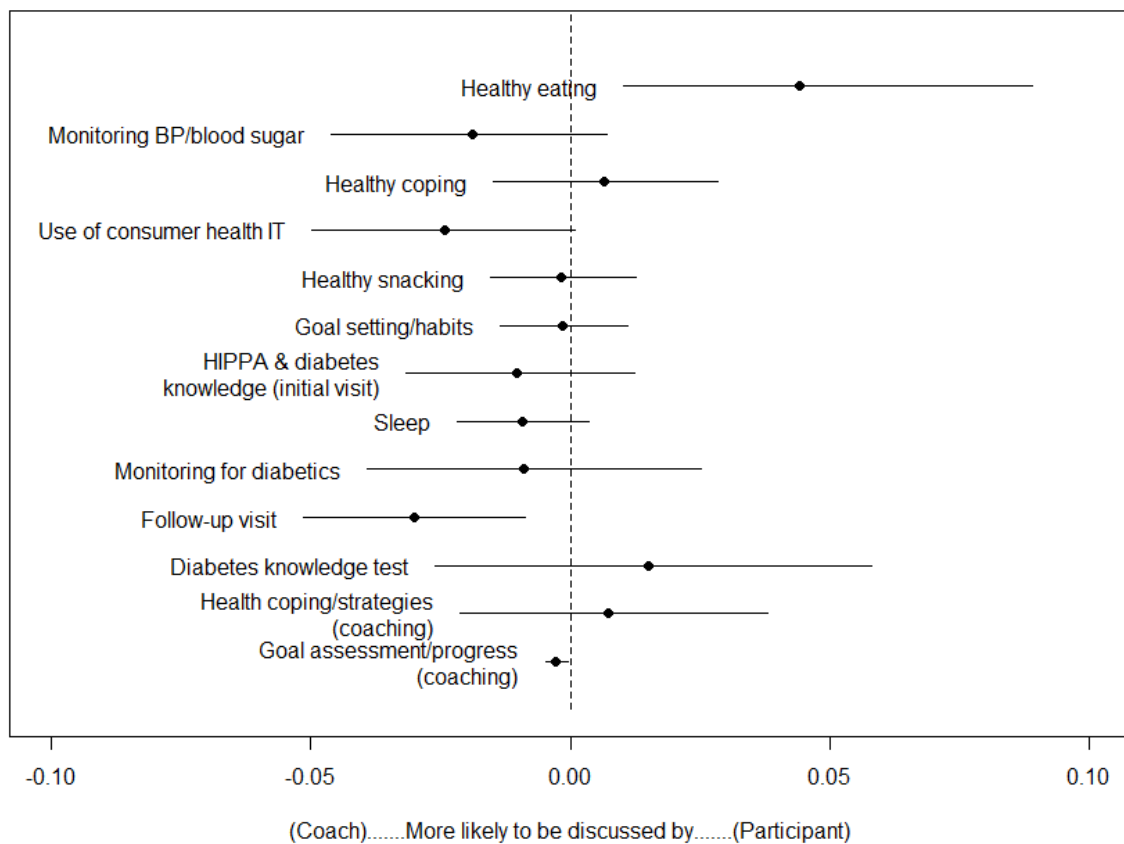
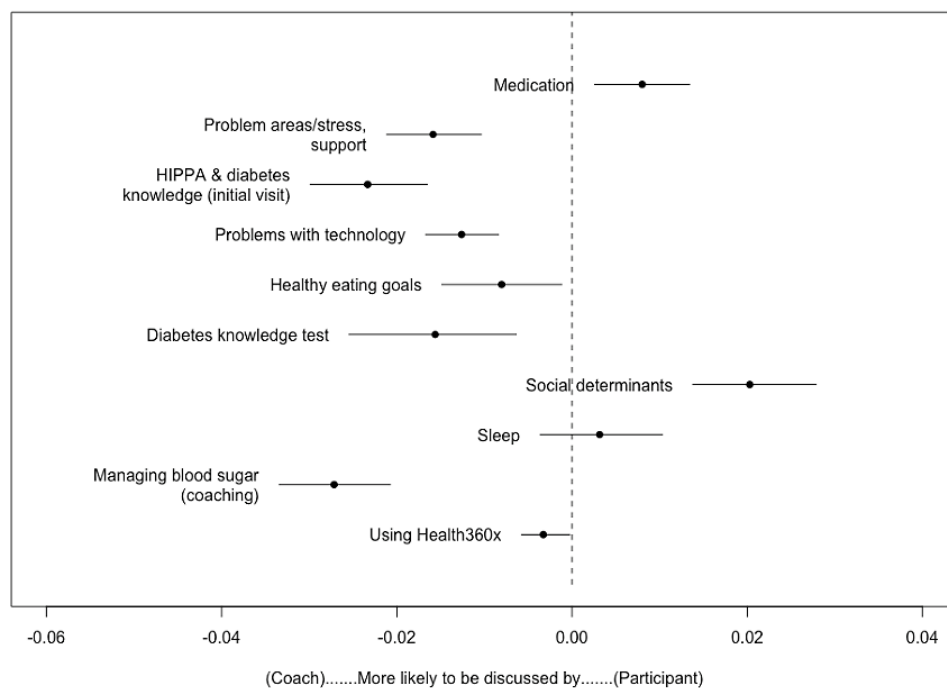


Figure 3. Mean difference in not achieved topic proportions—patients versus coaches. HIPAA: Health Insurance Portability and Accountability Act.



Discussion

Principal Findings

In this paper, we present the results of a novel analysis of the discursive emphasis on patients and coaches. The dialog is derived from an implementation of Health360x, a coach-assisted CHIT designed to support diabetes self-management skills acquisition and behavior for underserved, high-risk diabetic patients. Thematic analysis of conversations clustered using STM results of patient-coach interactions indicates that in patient-coach pairs that were able to achieve self-management goals, coaches provided more contextual feedback and probed into patients' familiarity with technology and use and trust of consumer information technologies. We also found that topics of conversation between successful patient-coach pairs were equally distributed, whereas discussions with patients who were unable to achieve enhanced self-efficacy for diabetes-related self-management were dominated by coaches.

Limitations

There are a few limitations of the analysis presented in this paper. First is the small corpus of transcribed conversations. As with all machine learning algorithms, the more training data, the better the models perform. Typically, topic models are used on corpora with tens of thousands of talk turns; however, we analyzed significantly less (approximately 17,000) talk turns. With that said, the methodological novelty of this paper is the incorporation of covariates into our estimation of topic models using the STM. Although we had a smaller than expected corpus, we found that we could reliably identify latent topics in comparatively smaller corpora by incorporating metadata into topic modeling. Transcription of the sessions was also outsourced and included errors, which may have further limited the overall information content of the transcripts.

A second potential limitation is that there are subjective decisions that go into creating a topic model that can influence topic outputs. The number of topics estimated as well as decisions around preprocessing (eg, removing stop words, stemming, and lemmatization) can influence topic model coherence and stability [31-33]. We addressed the subjectivity of arbitrarily selecting a number of topics by automating the number of topics to be estimated; however, another approach is to estimate and compare topic models with different topic sizes for stability across models. Finally, evaluations of preprocessing on topic model outputs are needed within the context of the STM used in this paper.

Another limitation relates to the lack of power in our analysis. The sample size calculation of 97 participants for this study was based on an effect size with a *P* value that is effectively .05. Moreover, our analysis of discursive analysis included transcripts from only half of the estimated sample size. To assess whether patient characteristics influence retention and transcript quality, we ran a logistic model regressing whether we had at least one session recording of high enough quality for analysis of patient characteristics. The results are presented in [Multimedia Appendix 4](#). Patient characteristics, including gender, age, insurance status, and willingness to use the internet, were not significantly different between patients whose session

recordings were retained and those whose session recordings were not.

A final limitation was related to the training of coaches, which was naturally limited. Coaches did not receive extensive training in cognitive behavioral therapy or motivational interviewing. Our training built the coaches' existing skills and provided additional tools to help with an appreciative inquiry. In the future, we will use these learnings to provide more uniform training for health coaches.

Comparison With Prior Work

A need exists to better understand how DSME might be delivered within primary care, the outcomes that can be achieved, and the training and system changes needed as a result [34]. Effective face-to-face multicomponent interventions require skilled primary care workers and activated patients coupled with tailored training and self-help systems, resources, and materials [35]. These interventions for diabetes self-management are complex, and their effectiveness is in large part determined by dynamics within the patient-provider relationship [36,37]. Where these relationships are 1-sided, that is, focus on what the patient *should* be doing, and do not account for a patient's circumstance, strategies for encouraging diabetes-related self-management behaviors will be ineffective [38].

Patients stand the best chance of developing skills to actively participate and take responsibility in the management of their chronic condition when they receive personalized coaching. Polonsky and Fisher [39] provided a conceptual model of how the frequency of feedback, personal meaningfulness, clarity, guidance and support, and patient characteristics influence diabetes-related feedback and affect patient engagement and behavior change. The way each of these elements should be personalized for diabetes-related coaching to be effective is beyond the scope of this comparison with prior work. What is relevant is that our analysis of the discursive emphasis on patients and coaches provides a novel approach for quantifying the substance of these relationships. Examples of coaches providing contextual feedback, probing patient experiences with technology, and engaging in back-and-forth conversations are all indicators of tailored coaching. However, what remains less clear is how we can relate these conversations to clinical outcomes and how we can quantify the additional benefit that technology provides in addition to tailored coaching. These 2 questions are the focus of subsequent projects and manuscripts.

Our study also contributes to prior work by providing new and important insights into the usability and effectiveness of mobile health interventions targeting older African American patients living with diabetes. This underserved population has traditionally been considered not amenable to the use of technology to manage their health [4]. In addition, we present novel methods that can be used to understand (and predict) what is happening in patient-coach interactions. These methods enable the analysis of themes in text at scale using techniques that minimize the amount of human-injected bias into qualitative studies.

Conclusions

Coach-facilitated, technology-based DSME can help address critical issues and unmet needs for underserved patients with diabetes around accessing sufficient support regularly. A variety of factors influence the efficacy of coaching; however, cultural expectations must be further explored, given the growing diversity among patients and the health care workforce prevalent

today. Our use of structural topic modeling in this application is novel, and it creates an opportunity to introduce this technique into everyday patient-provider encounters, particularly in a post-COVID-19 world in which there has been a paradigm shift around the use of telehealth. The opportunity to create outputs that guide further physician action and patient action could drive better patient engagement and overall patient health outcomes.

Authors' Contributions

PP and E Ofili conceptualized the project. PP and MYI contributed to the development of the project. E Olorundare, MM, and MB led the investigation. EAM conducted the data curation. MYI conducted the formal analysis. PP and E Ofili led the investigation. MYI contributed to writing the manuscript and preparing the original draft. PP and E Ofili completed the finding acquisition. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

E Ofili is the Founder and Chief Science Officer for AccuHealth Technologies, a technology startup which holds the patent for the Health360x platform. AccuHealth Technologies was not involved in the design and analysis of the study.

Multimedia Appendix 1

Example of contextual feedback in health coaching around health coping and strategies.

[[DOCX File, 23 KB - jmir_v24i3e23535_app1.docx](#)]

Multimedia Appendix 2

Example of consumer health information technology problems and use in the not achieved subgroup.

[[DOCX File, 20 KB - jmir_v24i3e23535_app2.docx](#)]

Multimedia Appendix 3

Example of probing questions around consumer health information technology use in the achieved subgroup.

[[DOCX File, 20 KB - jmir_v24i3e23535_app3.docx](#)]

Multimedia Appendix 4

Logistic regression of at least one recording on patient characteristics.

[[DOCX File, 13 KB - jmir_v24i3e23535_app4.docx](#)]

References

1. The social life of health information. Pew Internet & American Life Project. 2009. URL: https://web.archive.org/web/20091224000545/http://www.pewinternet.org/~media/Files/Reports/2009/PIP_Health_2009.pdf [accessed 2022-02-03]
2. Health topics: 80% of internet users look for health information online. The Communication Initiative Network. URL: <https://www.comminit.com/content/health-topics-80-internet-users-look-health-information-online> [accessed 2022-02-03]
3. Mobile Health 2010. Pew Research Center. URL: <https://www.pewresearch.org/internet/2010/10/19/mobile-health-2010/> [accessed 2022-02-03]
4. Christopher Gibbons M. Use of health information technology among racial and ethnic underserved communities. *Perspect Health Inf Manag* 2011 Jan 01;8:1f [FREE Full text] [Medline: 21307989]
5. Institute of Medicine, Board on Neuroscience and Behavioral Health, Committee on Health Literacy. In: Nielsen-Bohlman L, Panzer AM, Kindig DA, editors. *Health Literacy: A Prescription to End Confusion*. Washington, DC: The National Academies Press; 2004:41-42.
6. Jensen J, King A, Davis L, Guntzville L. Utilization of internet technology by low-income adults: the role of health literacy, health numeracy, and computer assistance. *J Aging Health* 2010 Sep;22(6):804-826. [doi: [10.1177/0898264310366161](https://doi.org/10.1177/0898264310366161)] [Medline: 20495159]
7. Sarkar U, Karter A, Liu J, Adler N, Nguyen R, Lopez A, et al. The literacy divide: health literacy and the use of an internet-based patient portal in an integrated health system-results from the diabetes study of northern California (DISTANCE). *J Health Commun* 2010;15 Suppl 2:183-196 [FREE Full text] [doi: [10.1080/10810730.2010.499988](https://doi.org/10.1080/10810730.2010.499988)] [Medline: 20845203]
8. Levine R, Rust G, Pisu M, Agboto V, Baltrus P, Briggs N, et al. Increased Black-White disparities in mortality after the introduction of lifesaving innovations: a possible consequence of US federal laws. *Am J Public Health* 2010 Nov;100(11):2176-2184 [FREE Full text] [doi: [10.2105/AJPH.2009.170795](https://doi.org/10.2105/AJPH.2009.170795)] [Medline: 20864727]

9. Toobert D, Strycker L, King D, Barrera M, Osuna D, Glasgow R. Long-term outcomes from a multiple-risk-factor diabetes trial for Latinas: ¡Viva Bien!. *Transl Behav Med* 2011 Sep;1(3):416-426 [FREE Full text] [doi: [10.1007/s13142-010-0011-1](https://doi.org/10.1007/s13142-010-0011-1)] [Medline: [22022345](https://pubmed.ncbi.nlm.nih.gov/22022345/)]
10. Steinsbekk A, Rygg L, Lisulo M, Rise MB, Fretheim A. Group based diabetes self-management education compared to routine treatment for people with type 2 diabetes mellitus. A systematic review with meta-analysis. *BMC Health Serv Res* 2012 Jul 23;12:213 [FREE Full text] [doi: [10.1186/1472-6963-12-213](https://doi.org/10.1186/1472-6963-12-213)] [Medline: [22824531](https://pubmed.ncbi.nlm.nih.gov/22824531/)]
11. Deakin T, McShane CE, Cade JE, Williams RD. Group based training for self-management strategies in people with type 2 diabetes mellitus. *Cochrane Database Syst Rev* 2005 Apr 18(2):CD003417. [doi: [10.1002/14651858.CD003417.pub2](https://doi.org/10.1002/14651858.CD003417.pub2)] [Medline: [15846663](https://pubmed.ncbi.nlm.nih.gov/15846663/)]
12. Gary TL, Genkinger JM, Guallar E, Peyrot M, Brancati FL. Meta-analysis of randomized educational and behavioral interventions in type 2 diabetes. *Diabetes Educ* 2003;29(3):488-501. [doi: [10.1177/014572170302900313](https://doi.org/10.1177/014572170302900313)] [Medline: [12854339](https://pubmed.ncbi.nlm.nih.gov/12854339/)]
13. Cochran J, Conn VS. Meta-analysis of quality of life outcomes following diabetes self-management training. *Diabetes Educ* 2008;34(5):815-823 [FREE Full text] [doi: [10.1177/0145721708323640](https://doi.org/10.1177/0145721708323640)] [Medline: [18832286](https://pubmed.ncbi.nlm.nih.gov/18832286/)]
14. Sturt J, Whitlock S, Fox C, Hearnshaw H, Farmer A, Wakelin M, et al. Effects of the Diabetes Manual 1:1 structured education in primary care. *Diabet Med* 2008 Jun;25(6):722-731. [doi: [10.1111/j.1464-5491.2008.02451.x](https://doi.org/10.1111/j.1464-5491.2008.02451.x)] [Medline: [18435777](https://pubmed.ncbi.nlm.nih.gov/18435777/)]
15. Duncan I, Birkmeyer C, Coughlin S, Li Q, Sherr D, Boren S. Assessing the value of diabetes education. *Diabetes Educ* 2009;35(5):752-760. [doi: [10.1177/0145721709343609](https://doi.org/10.1177/0145721709343609)] [Medline: [19783766](https://pubmed.ncbi.nlm.nih.gov/19783766/)]
16. Powers M, Bardsley J, Cypress M, Duker P, Funnell M, Fischl AH, et al. Diabetes self-management education and support in type 2 diabetes: a joint position statement of the american diabetes association, the american association of diabetes educators, and the academy of nutrition and dietetics. *J Acad Nutr Diet* 2015 Aug;115(8):1323-1334. [doi: [10.1016/j.jand.2015.05.012](https://doi.org/10.1016/j.jand.2015.05.012)] [Medline: [26054423](https://pubmed.ncbi.nlm.nih.gov/26054423/)]
17. Couper M, Alexander G, Zhang N, Little R, Maddy N, Nowak M, et al. Engagement and retention: measuring breadth and depth of participant use of an online intervention. *J Med Internet Res* 2010 Nov 18;12(4):e52 [FREE Full text] [doi: [10.2196/jmir.1430](https://doi.org/10.2196/jmir.1430)] [Medline: [21087922](https://pubmed.ncbi.nlm.nih.gov/21087922/)]
18. Tang T, Funnell M, Oh M. Lasting effects of a 2-year diabetes self-management support intervention: outcomes at 1-year follow-up. *Prev Chronic Dis* 2012;9:E109 [FREE Full text] [doi: [10.5888/pcd9.110313](https://doi.org/10.5888/pcd9.110313)] [Medline: [22677159](https://pubmed.ncbi.nlm.nih.gov/22677159/)]
19. Pemu P, Josiah Willock R, Alema-Mensa E, Rollins L, Brown M, Saint Clair B, et al. Achieving health equity with e-Healthystrides©: patient perspectives of a consumer health information technology application. *Ethn Dis* 2019;29(Suppl 2):393-404 [FREE Full text] [doi: [10.18865/ed.29.S2.393](https://doi.org/10.18865/ed.29.S2.393)] [Medline: [31308611](https://pubmed.ncbi.nlm.nih.gov/31308611/)]
20. Heisler M, Choi H, Mase R, Long JA, Reeves PJ. Effectiveness of technologically enhanced peer support in improving glycemic management among predominantly African American, low-income adults with diabetes. *Diabetes Educ* 2019 Jun;45(3):260-271 [FREE Full text] [doi: [10.1177/0145721719844547](https://doi.org/10.1177/0145721719844547)] [Medline: [31027477](https://pubmed.ncbi.nlm.nih.gov/31027477/)]
21. Ofili E, Pemu P, Quarshie A, Johnson P, Linn L, Kaur D, et al. System and method for chronic illness care. Science On. URL: <https://scienceon.kisti.re.kr/srch/selectPORSrchPatent.do?cn=USP2012078234131> [accessed 2022-02-03]
22. Brown CH, Curran G, Palinkas LA, Aarons GA, Wells KB, Jones L, et al. An overview of research and evaluation designs for dissemination and implementation. *Annu Rev Public Health* 2017 Mar 20;38:1-22 [FREE Full text] [doi: [10.1146/annurev-publhealth-031816-044215](https://doi.org/10.1146/annurev-publhealth-031816-044215)] [Medline: [28384085](https://pubmed.ncbi.nlm.nih.gov/28384085/)]
23. Brown M, Ofili EO, Okirie D, Pemu P, Franklin C, Suk Y, et al. Morehouse Choice Accountable Care Organization and Education System (MCACO-ES): integrated model delivering equitable quality care. *Int J Environ Res Public Health* 2019 Aug 25;16(17):3084 [FREE Full text] [doi: [10.3390/ijerph16173084](https://doi.org/10.3390/ijerph16173084)] [Medline: [31450652](https://pubmed.ncbi.nlm.nih.gov/31450652/)]
24. Webb T, 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](https://doi.org/10.2196/jmir.1376)] [Medline: [20164043](https://pubmed.ncbi.nlm.nih.gov/20164043/)]
25. Howes C, Purver M, McCabe R. Using conversation topics for predicting therapy outcomes in schizophrenia. *Biomed Inform Insights* 2013 Jul 15;6(Suppl 1):39-50 [FREE Full text] [doi: [10.4137/BII.S11661](https://doi.org/10.4137/BII.S11661)] [Medline: [23943658](https://pubmed.ncbi.nlm.nih.gov/23943658/)]
26. Park J, Kotzias D, Kuo P, Logan Iv RL, Merced K, Singh S, et al. Detecting conversation topics in primary care office visits from transcripts of patient-provider interactions. *J Am Med Inform Assoc* 2019 Dec 01;26(12):1493-1504 [FREE Full text] [doi: [10.1093/jamia/ocz140](https://doi.org/10.1093/jamia/ocz140)] [Medline: [31532490](https://pubmed.ncbi.nlm.nih.gov/31532490/)]
27. 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 Polit Sci* 2014 Mar 06;58(4):1064-1082. [doi: [10.1111/ajps.12103](https://doi.org/10.1111/ajps.12103)]
28. Blei D, Ng A, Jordan MI. Latent dirichlet allocation. *J Mach Learn Res* 2003 Jan 3;3:993-1022. [doi: [10.5555/944919.944937](https://doi.org/10.5555/944919.944937)]
29. Blei DM, Lafferty JD. A correlated topic model of science. *Annals Appl Stat* 2007 Jun 1;1(1):17-35 [FREE Full text] [doi: [10.1214/07-aos114](https://doi.org/10.1214/07-aos114)]
30. Mimno D, Lee M. Low-dimensional embeddings for interpretable anchor-based topic inference. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Oct 25-29, 2014; Doha, Qatar. [doi: [10.3115/v1/d14-1138](https://doi.org/10.3115/v1/d14-1138)]

31. Denny M, Spirling A. Text Preprocessing For Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It. *Political Analysis* 2018;26(2):168-189. [doi: [10.1017/pan.2017.44](https://doi.org/10.1017/pan.2017.44)]
32. Schofield A, Mimno D. Comparing apples to apple: the effects of stemmers on topic models. *Transact Assoc Comput Linguistics* 2016 Dec;4(2):287-300. [doi: [10.1162/tacl_a_00099](https://doi.org/10.1162/tacl_a_00099)]
33. Understanding text pre-processing for latent dirichlet allocation. Cornell. URL: <https://www.cs.cornell.edu/~xanda/winlp2017.pdf> [accessed 2022-02-03]
34. Dineen-Griffin S, Garcia-Cardenas V, Williams K, Benrimoj S. Helping patients help themselves: a systematic review of self-management support strategies in primary health care practice. *PLoS One* 2019;14(8):e0220116 [FREE Full text] [doi: [10.1371/journal.pone.0220116](https://doi.org/10.1371/journal.pone.0220116)] [Medline: [31369582](https://pubmed.ncbi.nlm.nih.gov/31369582/)]
35. Jonkman NH, Schuurmans MJ, Jaarsma T, Shortridge-Baggett LM, Hoes AW, Trappenburg JC. Self-management interventions: proposal and validation of a new operational definition. *J Clin Epidemiol* 2016 Dec;80:34-42. [doi: [10.1016/j.jclinepi.2016.08.001](https://doi.org/10.1016/j.jclinepi.2016.08.001)] [Medline: [27531245](https://pubmed.ncbi.nlm.nih.gov/27531245/)]
36. Schillinger D, Piette J, Grumbach K, Wang F, Wilson C, Daher C, et al. Closing the loop: physician communication with diabetic patients who have low health literacy. *Arch Intern Med* 2003 Jan 13;163(1):83-90. [doi: [10.1001/archinte.163.1.83](https://doi.org/10.1001/archinte.163.1.83)] [Medline: [12523921](https://pubmed.ncbi.nlm.nih.gov/12523921/)]
37. O'Connor PJ, Sperl-Hillen J, Johnson PE, Rush WA, Crain AL. Customized feedback to patients and providers failed to improve safety or quality of diabetes care: a randomized trial. *Diabetes Care* 2009 Jul;32(7):1158-1163 [FREE Full text] [doi: [10.2337/dc08-2247](https://doi.org/10.2337/dc08-2247)] [Medline: [19366977](https://pubmed.ncbi.nlm.nih.gov/19366977/)]
38. Hood KK, Hilliard M, Piatt G, Ievers-Landis CE. Effective strategies for encouraging behavior change in people with diabetes. *Diabetes Manag (Lond)* 2015;5(6):499-510 [FREE Full text] [Medline: [30100925](https://pubmed.ncbi.nlm.nih.gov/30100925/)]
39. Polonsky WH, Fisher L. When does personalized feedback make a difference? A narrative review of recent findings and their implications for promoting better diabetes self-care. *Curr Diab Rep* 2015 Aug;15(8):50. [doi: [10.1007/s11892-015-0620-7](https://doi.org/10.1007/s11892-015-0620-7)] [Medline: [26077015](https://pubmed.ncbi.nlm.nih.gov/26077015/)]

Abbreviations

CHIT: consumer health information technology

DSME: diabetes self-management education

MCACO: Morehouse Choice Accountable Care Organization

STM: structural topic model

Edited by C Richardson; submitted 16.08.20; peer-reviewed by J Luo, Z Ren, A Alasmari, I Shubina, R Krukowski; comments to author 18.11.20; revised version received 13.01.21; accepted 19.11.21; published 18.03.22.

Please cite as:

Idris MY, Alema-Mensah E, Olorundare E, Mohammad M, Brown M, Ofili E, Pemu P

Exploring the Discursive Emphasis on Patients and Coaches Who Participated in Technology-Assisted Diabetes Self-management Education: Clinical Implementation Study of Health360x

J Med Internet Res 2022;24(3):e23535

URL: <https://www.jmir.org/2022/3/e23535>

doi: [10.2196/23535](https://doi.org/10.2196/23535)

PMID: [35302506](https://pubmed.ncbi.nlm.nih.gov/35302506/)

©Muhammed Y Idris, Ernest Alema-Mensah, Elizabeth Olorundare, Mohammad Mohammad, Michelle Brown, Elizabeth Ofili, Priscilla Pemu. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 18.03.2022. 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.

Original Paper

Modeling Access Across the Digital Divide for Intersectional Groups Seeking Web-Based Health Information: National Survey

Kristina Medero¹, MA; Kelly Merrill Jr¹, MA; Morgan Quinn Ross¹, BA

School of Communication, Ohio State University, Columbus, OH, United States

Corresponding Author:

Kristina Medero, MA
School of Communication
Ohio State University
154 N Oval Mall
Columbus, OH, 43210-1132
United States
Phone: 1 5313017556
Email: medero.2@osu.edu

Abstract

Background: The digital divide refers to technological disparities based on demographic characteristics (eg, race and ethnicity). Lack of physical access to the internet inhibits online health information seeking (OHIS) and exacerbates health disparities. Research on the digital divide examines where and how people access the internet, whereas research on OHIS investigates how intersectional identities influence OHIS. We combine these perspectives to explicate how unique context–device access pairings operate differently across intersectional identities—particularly racial and ethnic groups—in the domain of OHIS.

Objective: This study aims to examine how different types of internet access relate to OHIS for different racial and ethnic groups. We investigate relationships among predisposing characteristics (ie, age, sex, education, and income), internet access (home computer, public computer, work computer, and mobile), health needs, and OHIS.

Methods: Analysis was conducted using data from the 2019 Health Information National Trends Survey. Our theoretical model of OHIS explicates the roles of internet access and health needs for racial and ethnic minority groups' OHIS. Participant responses were analyzed using structural equation modeling. Three separate group structural equation modeling models were specified based on Black, Latine, and White self-categorizations.

Results: Overall, predisposing characteristics (ie, age, sex, education, and income) were associated with internet access, health needs, and OHIS; internet access was associated with OHIS; and health needs were associated with OHIS. Home computer and mobile access were most consistently associated with OHIS. Several notable linkages between predisposing characteristics and internet access differed for Black and Latine individuals. Older racial and ethnic minorities tended to access the internet on home and public computers less frequently; home computer access was a stronger predictor of OHIS for White individuals, and mobile access was a stronger predictor of OHIS for non-White individuals.

Conclusions: Our findings necessitate a deeper unpacking of how physical internet access, the foundational and multifaceted level of the digital divide, affects specific racial and ethnic groups and their OHIS. We not only find support for prior work on the digital divide but also surface new insights, including distinct impacts of context–device access pairings for OHIS and several relationships that differ between racial and ethnic groups. As such, we propose interventions with an intersectional approach to access to ameliorate the impact of the digital divide.

(*J Med Internet Res* 2022;24(3):e32678) doi:[10.2196/32678](https://doi.org/10.2196/32678)

KEYWORDS

Black; African American; first-level digital divide; health disparities; home computer; internet access; intersectionality; Latino; Latine; Hispanic; mobile; online health information seeking; public computer; structural equation modeling; work computer; mobile phone

Introduction

Background

The benefits of eHealth, or the use of the internet to facilitate health behaviors (eg, online health information seeking [OHIS]) [1], are counteracted by the digital divide. The digital divide was first used to emphasize that racial and ethnic minorities and individuals of lower socioeconomic status did not adopt new technologies to the same extent as White individuals or those of higher socioeconomic status [2]. Obtaining physical access to new technologies and, thus, web-based health information remains a paramount obstacle, particularly for Black or African American (hereafter *Black*) and Latino or Latine or Hispanic (hereafter *Latine*) individuals [3]. This is problematic, as racial and ethnic minorities are more likely to live in areas of concentrated poverty that coincide with limited health care access [4-6]. Systematic inequalities in internet access and health care for racial and ethnic minorities reinforce one another, such that those who would potentially benefit the most from OHIS are often unable to access it.

However, internet access (hereafter *access*) is not monolithic and comprises the use of different devices (eg, smartphone or computer) in various contexts (eg, at home or in public) [7]. Although mobile devices are increasingly more accessible than computers, they can be harder to navigate because of their smaller interfaces [8]. Publicly accessible devices may extend access to those who do not own such devices; however, they often entail irregular availability, which can compound poor health outcomes for minority groups [9]. Recognizing the multidimensionality of access [7] is key to understanding how access via a myriad of devices in various contexts differentially influences OHIS. Furthermore, positioning the digital divide as a health disparity is imperative to developing effective interventions [10]. As such, this study uses data from the 2019 Health Information National Trends Survey (HINTS), also known as HINTS 5, Cycle 3 [11], to bolster theoretical models of OHIS with a nuanced conceptualization of access. We also advance the perspective that the digital divide is a health disparity by applying an intersectional focus to examine how relationships with access and OHIS differ across racial and ethnic groups.

Theoretical Framework: OHIS Model

The internet has become one of the most common ways of accessing health information [12]. Health information seeking refers to those actions that individuals use to search for information about their health, risks, illnesses, and health-protective behaviors [13]. When conducted on the web (ie, OHIS), seeking out health information can positively affect health outcomes by improving the quality, expense, and efficiency of health care [10]. In addition, OHIS has demonstrated that individuals are more willing to comply with their health decisions [14]. However, those with limited access to OHIS may not experience its benefits. Health disparities faced by low-income and minority communities may be magnified by the digital divide [3,15]. However, when underserved communities are provided the means to participate in OHIS, they gain more health knowledge [16]. Thus,

understanding how the digital divide affects OHIS is imperative to enhance the impact of interventions aimed at increasing access among these communities.

The digital divide first highlighted that certain groups of people (eg, racial and ethnic minorities and individuals of low socioeconomic status) lagged in adopting new technologies. This gradual diffusion represents the first-level factor of the digital divide, which has been situated in issues related to ownership, availability, and affordability of the technology [17]. Recent studies have identified additional second-level factors that may also impede technological adoption (eg, skills) [15,18]. Although the focus of OHIS has shifted away from access as some suggest that it has become democratized [19], we argue that it has not been democratized across devices and contexts of use as the lack of physical access remains an obstacle for marginalized groups [3,20]. Moreover, access is heterogeneous, as people can access the internet on multiple devices and at various places [7,21]. Even in populations with saturated home access, disparities can persist for other points of access and the cost to maintain them [7]. Thus, a nuanced conceptualization of access can respond to criticism that the digital divide suggests a simple binary between those who have access and those who do not [22].

Notably, some scholars have applied this multifaceted conceptualization of access to predict the likelihood of web-based activities (including OHIS). Hassani [23] found that people engaged in more OHIS as they increased their points of access (eg, home and work vs only home). Similarly, Mossberger et al [24] found that home computer access is vital to reap the benefits of web-based activities such as OHIS. Although the authors highlight the potential for mobile devices to attenuate the impacts of the digital divide, mobile access alone did little to minimize these impacts in low-income areas. Reisdorf et al [21] also suggest that simply increasing access does not lead to equal results across different contexts. However, these studies investigated OHIS as one of many web-based activities; as such, they were not grounded in theoretical models of OHIS. Moreover, studies that focused on OHIS [25,26] did not examine the impact of specific devices or contexts of use; instead, they examined the number of access points overall.

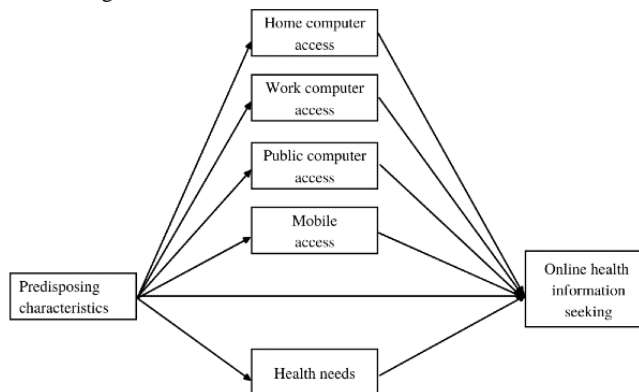
Furthermore, scholarship in this area has seldom disaggregated these connections by racial and ethnic groups. Studies that include race and ethnicity self-categorization as predictors of web-based activities [21] can unearth patterns of devices and contexts of use, such as Black (vs non-Black) individuals using mobile devices more often [24]. However, disparities among intersectional identities may still be overlooked, such as how age and race may interact to affect technology use [27]. Therefore, it is unclear where disparities in OHIS exist among intersectional identities.

Previous OHIS theorizing [14,28] highlights several factors that influence health-seeking behaviors. As such, our model includes predisposing characteristics (age, sex, education, and income), access (home computer, work computer, public computer, and mobile), health needs, and OHIS (Figure 1). Our model also focuses on the foundational level of the digital divide (ie, access). As such, our nuanced conceptualization of access helps

to fill the empirical gap in OHIS research on the first-level digital divide [29]. Thus, this study offers 2 primary contributions. First, we apply a multidimensional conceptualization of access along dimensions of the context of use (eg, at home vs in public) and device type (eg, smartphone

vs computer) to theoretical models of OHIS. Second, we disaggregate our models by race and ethnicity to examine how they manifest differently across racial and ethnic groups. We explicate the relationships between predisposing characteristics, access, and health needs in our model.

Figure 1. Initial online health information seeking model.



Antecedents of OHIS

Predisposing Characteristics

First, we posit associations between OHIS and predisposing characteristics. Younger individuals are more likely than older individuals to be able to navigate web-based platforms to seek out web-based health information [30,31]. Females are often expected to seek out health information because of their social roles as family caregivers [23,32]. Furthermore, education and income are generally positively associated with OHIS [33,34], as individuals with low education or income are inhibited from participating in OHIS because of low literacy [5]. Thus, we propose the following hypothesis:

Hypothesis 1 (H1): Demographic variables—age, sex, education, and income—will be associated with OHIS.

Internet Access

Expanding our understanding of OHIS, we extend prior conceptualizations of access [7,21,23] and examine access by considering context and device. *Context* refers to the physical environment in which users engage in OHIS, and *device* refers to the physical technology used to engage in OHIS. We focus on four of the most common context–device pairings: home computer, work computer, public computer, and mobile (which can be used across contexts). People seek health information on the web on several devices and at several places [23]. Owing to the extent that computers and mobile devices entail different technological constraints [8], and context structures media use [35], it is crucial to consider how different context–device pairings relate to OHIS.

First, *home computer* access involves computer use at home. It facilitates the availability of OHIS in a private setting where people spend most of their time [23]. However, because of the large cost of computers, ownership trends reflect the digital divide: White individuals are more likely to own home computers than Black and Latine individuals [3,7].

Second, *work computer* access involves computer use in the workplace. Individuals may not own these devices and likely would not pay for access, making it less expensive than home-computer access. However, work computer access requires employment that involves or entails access to computers [36]. The workplace is also a less frequented and private setting than the home [23,37].

Third, *public computer* access involves computer use in public facilities [3]. Such access can be inexpensive (if not free) and occurs in typically accessible public places, enabling access for individuals who cannot afford devices with internet access [21]. However, public computer access is contingent on a variety of factors, including the hours, locations, and resources of public facilities, which restrict the availability of such services [5].

Finally, *mobile* access involves the use of mobile devices (typically smartphones) in a variety of contexts. Mobile devices allow users to connect to public Wi-Fi and data networks, enabling OHIS in a variety of public, private, and (uniquely) mobile places (eg, on the bus) [38]. However, the small size of the mobile interface may restrict more intensive tasks [8], such as OHIS.

Overall, we expect that access will be related to predisposing characteristics and OHIS. Older individuals are less likely to access the internet [7,39], and males tend to access the internet more than females [18]. Education and income are positively correlated with access [40,41]. These differences may be linked to literacy and resources, allowing certain groups to maintain [7] and navigate access [10]. Furthermore, OHIS, by definition, is contingent on internet access [42,43]. With these considerations in mind, we propose the following hypotheses:

Hypothesis 2 (H2): Predisposing characteristics (ie, age, sex, education, and income) will be associated with access.

Hypothesis 3 (H3): Access will be positively associated with OHIS.

However, it is unclear how our nuanced conceptualization of access (ie, 4 discrete context–device pairings) may differentially

affect OHIS. Thus, we pose the following research question (RQ):

RQ1: Which access pairings have the most consistent associations with OHIS across racial and ethnic groups?

Health Need

We conceptualize health needs as the extent to which individuals perceive that they require current or chronic medical attention. The likelihood that one may endure chronic illness is linked to group identities along the lines of age, gender, education level, and income [44]. When avoidable health differences in treatment, access to treatment, mortality, and diseases correlate with group identity, a health disparity occurs. Older individuals report greater health needs than younger individuals [45]. Although health disparities among males and females differ based on the illness, males may be more confident in their ability to maintain their health and report lower health needs [46]. Braveman et al [44] highlight that education level and income are important health determinants that predict health needs. Finally, individuals who perceive their health to be poor often demonstrate motivation to find health information on the web [28,33]. Thus, we propose the following hypotheses:

Hypothesis 4 (H4): Predisposing characteristics (ie, age, sex, education, and income) will be associated with health needs.

Hypothesis 5 (H5): Health needs will be positively associated with OHIS.

Race and Ethnicity

This study holds that existing racial and ethnic disparities exacerbate the impact of the digital divide on health disparities [22]. As such, we investigate how these inequities may affect OHIS. We explore whether hypotheses linking predisposing characteristics with OHIS (H1), access (H2), and health need (H4) differ across racial and ethnic groups. First, race and ethnicity may interact with age, sex, education, and income to predict OHIS as unique disparities in health and technology have been observed within groups that have intersecting predisposing characteristics and racial and ethnic group identities [41]. Next, race and ethnicity may interact with access to predict OHIS. Fang et al [47] provide illustrative insights, highlighting that age was a strong predictor for access but that this effect was exaggerated for some racial and ethnic minorities. Finally, race and ethnicity may interact with health needs to predict OHIS. For example, Black and Latine individuals are more likely to live in low-income areas [4,6], which is associated with exacerbated health needs [48]. As such, this study foregrounds the persistent racial and ethnic disparities in the United States to understand access and health needs from an intersectionality perspective [49].

In addition, race and ethnicity may interact with access (H3) and health need (H5) to influence OHIS. Regarding access, even when Black and Latine individuals access the internet at similar rates as White individuals, such access is often marked by greater insecurity [39]. Similarly, even with similar levels of health needs, racial and ethnic minorities may avoid seeking out web-based health information if they possess lower health

and technology literacy [50,51]. Overall, our study uses previously tested models of OHIS [14,28,50-52] and seeks to extend previous theories by applying a multidimensional conceptualization of access and testing the fit of the model across racial and ethnic groups. These intersectional considerations, heightened by higher rates of internet insecurity [39] and greater health needs [4,6] because of systemic inequality, beg the following question:

RQ2: How will the relationships between predisposing characteristics, access, health needs, and OHIS differ across different racial and ethnic groups?

Methods

Sample

To test our model, we used data from the 2019 HINTS, also known as HINTS 5, Cycle 3 [11]. HINTS is an annual, nationally representative survey that asks participants about their engagement with health information. Data were collected between January 2019 and April 2019. A total of 5438 individuals responded to the survey. However, of the 5438 responses, 191 (3.51%) responses were deemed ineligible by HINTS because of partial completion, leaving 5247 (96.49%) individuals. Participants who did not complete the self-categorization variables for each model were excluded (Black and White: 420/5247, 8%; Latine: 487/5247, 9.28%). In addition, participants who did not complete all model variables were excluded (Black and White: 446/5247, 8.12%; Latine: 408/5247, 7.78%). Taken together, the sample size for the final models' group was 4381 (based on Black and White self-categorization) or 4352 (based on Latine self-categorization). Owing to these different sample sizes, we report demographics and correlations for the 5247 individuals deemed eligible by HINTS. Only the variables presented in the *Measures* section were used for the purposes of this study; all other variables were excluded. Data are available in [Multimedia Appendix 1](#). More information regarding the methodology can be found in the 2019 HINTS methodology report [53].

Ethical Considerations

An institutional review board approval was not requested because the analysis for this study was conducted using secondary data. All HINTS data sets, including the one used for analysis in this study, have been approved through expedited review by the Westat Institutional Review Board, and subsequently deemed exempt by the U.S. National Institutes of Health Office of Human Subjects Research Protections [54].

Participant Demographics

Demographic data were used to assess predisposing factors. Participants were aged 56.58 (SD 16.88) years on average. Approximately 56.62% (2971/5247) of the participants self-categorized as female, and 41.16% (2160/5247) self-categorized as male. Race and ethnicity were operationalized in comparison with those who did not self-categorize as the respective racial or ethnic group as individuals who self-categorize ethnically as Latine may still self-categorize racially as White or Black. Of the 5247 individual, 3727 (71.03%) self-categorized as White, and 1100

(20.96%) did not; 847 (16.14%) self-categorized as Black and 3980 (75.85%) did not; and 716 (13.64%) participants self-categorized as Latine and 4044 (77.07%) did not. The remaining individuals did not disclose their sex, race, or ethnicity. Participants' level of education was measured on a

5-point scale from *less than high school* (score=1) to *postbaccalaureate degree* (score=5), and participants' annual income was measured on a 7-point scale from *US\$0* to *US \$19,999* (score=1) to \geq *US \$200,000* (score=7). See [Table 1](#) for a summary of participant demographics.

Table 1. Participant demographics.

Demographics	OHIS ^{a,b}	
	No, n (%)	Yes, n (%)
Age (years; n=5090)		
18-24	20 (0.39)	132 (2.59)
25-35	62 (1.22)	535 (10.51)
36-44	66 (1.30)	492 (9.67)
45-54	172 (3.38)	627 (12.32)
55-64	316 (6.21)	827 (16.25)
\geq 65	790 (15.52)	1051 (20.65)
Sex (n=5110)		
Male	652 (12.76)	1501 (29.37)
Female	798 (15.62)	2159 (42.25)
White (n=4805)		
No	344 (7.16)	746 (15.53)
Yes	977 (20.33)	2738 (56.98)
Black (n=4805)		
No	1034 (21.52)	2934 (61.06)
Yes	287 (5.97)	550 (11.45)
Latine (n=4745)		
No	1015 (21.39)	3016 (63.56)
Yes	237 (4.99)	477 (11.45)
Education (n=5087)		
Less than high school	200 (3.93)	108 (2.12)
High school graduate	445 (8.75)	448 (8.81)
Some college	441 (8.67)	1093 (21.49)
Received a bachelor's degree	230 (4.52)	1130 (22.21)
Received a postbaccalaureate degree	117 (2.30)	875 (17.2)
Income (US \$; n=4637)		
0-19,999	411 (8.86)	441 (9.51)
20,000-34,999	213 (4.59)	380 (8.19)
35,000-49,999	173 (3.73)	433 (9.34)
50,000-74,999	182 (3.92)	639 (13.78)
75,000-99,999	116 (2.5)	461 (9.94)
100,000-199,999	106 (2.29)	764 (16.48)
>200,000	36 (0.78)	282 (6.08)

^aOHIS: online health information seeking.

^bPercentages reflect those who responded to the OHIS item.

Measures**Overview**

Correlations between all variables are displayed in [Tables 2-4](#).

Table 2. Descriptive statistics and correlations between study variables (age, sex, and White).

Predictors	Values, mean (SD)	Age		Sex		White	
		<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value
Age (years)	56.58 (16.88)	1	— ^a	—	—	—	—
Sex ^b	0.42 (0.49)	0.04	.01	1	—	—	—
White ^c	0.77 (0.42)	0.04	.005	0.07	<.001	1	—
Black ^d	0.18 (0.38)	-0.01	.64	-0.09	<.001	-0.79	<.001
Latine ^e	0.15 (0.36)	-0.10	<.001	0.01	.68	0.10	<.001
Education	3.36 (1.16)	-0.17	<.001	0.03	.07	0.08	<.001
Income	3.76 (1.93)	-0.17	<.001	0.12	<.001	0.14	<.001
Home computer	1.15 (0.84)	-0.17	<.001	0.10	<.001	0.15	<.001
Work computer	0.70 (0.90)	-0.45	<.001	0.03	.06	0.06	<.001
Public computer	0.16 (0.39)	-0.20	<.001	0.01	.69	-0.07	<.001
Mobile	1.27 (0.85)	-0.51	<.001	-0.04	.004	0.08	<.001
Health needs	2.58 (0.94)	0.16	<.001	-0.02	.29	-0.09	<.001
Online health information seeking (OHIS)	0.71 (0.45)	-0.31	<.001	-0.04	.01	0.05	.001

^aNot applicable.

^bCoded as female=0 and male=1.

^cCoded as non-White=0 and White=1.

^dCoded as non-Black=0 and Black=1.

^eCoded as non-Latine=0 and Latine=1.

Table 3. Descriptive statistics and correlations between study variables (Black, Latine, education, and income).

Predictors	Black		Latine		Education		Income	
	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value
Age (years)	— ^a	—	—	—	—	—	—	—
Sex ^b	—	—	—	—	—	—	—	—
White ^c	—	—	—	—	—	—	—	—
Black ^d	1	—	—	—	—	—	—	—
Latine ^e	-0.10	<.001	1	—	—	—	—	—
Education	-0.12	<.001	-0.17	<.001	1	—	—	—
Income	-0.20	<.001	-0.12	<.001	0.47	<.001	1	—
Home computer	-0.15	<.001	-0.16	<.001	0.41	<.001	0.38	<.001
Work computer	-0.09	<.001	-0.08	<.001	0.40	<.001	0.46	<.001
Public computer	0.07	<.001	-0.01	.76	0.11	<.001	-0.03	.03
Mobile	-0.08	<.001	-0.04	.005	0.33	<.001	0.37	<.001
Health need	0.10	<.001	0.07	<.001	-0.25	<.001	-0.31	<.001
Online health information seeking (OHIS)	-0.07	<.001	-0.07	<.001	0.34	<.001	0.28	<.001

^aNot applicable.

^bCoded as female=0 and male=1.

^cCoded as non-White=0 and White=1.

^dCoded as non-Black=0 and Black=1.

^eCoded as non-Latine=0 and Latine=1.

Table 4. Descriptive statistics and correlations between study variables (home computer, work computer, public computer, mobile, and health need).

Predictors	Home computer		Work computer		Public computer		Mobile		Health need	
	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value	<i>r</i>	<i>P</i> value
Age (years)	— ^a	—	—	—	—	—	—	—	—	—
Sex ^b	—	—	—	—	—	—	—	—	—	—
White ^c	—	—	—	—	—	—	—	—	—	—
Black ^d	—	—	—	—	—	—	—	—	—	—
Latine ^e	—	—	—	—	—	—	—	—	—	—
Education	—	—	—	—	—	—	—	—	—	—
Income	—	—	—	—	—	—	—	—	—	—
Home computer	1	—	—	—	—	—	—	—	—	—
Work computer	0.38	<.001	1	—	—	—	—	—	—	—
Public computer	0.15	<.001	0.13	<.001	1	—	—	—	—	—
Mobile	0.48	<.001	0.48	<.001	0.21	<.001	1	—	—	—
Health need	-0.18	<.001	-0.23	<.001	-0.02	.14	-0.20	<.001	1	—
Online health information seeking (OHIS)	0.41	<.001	0.30	<.001	0.14	<.001	0.46	<.001	-0.10	<.001

^aNot applicable.

^bCoded as female=0 and male=1.

^cCoded as non-White=0 and White=1.

^dCoded as non-Black=0 and Black=1.

^eCoded as non-Latine=0 and Latine=1.

Internet Access

Participants reported how often they access the internet on a computer at home, at work, in a public place, and on a mobile device. A single item was used to measure each mode of access. Items were measured on 3-point scales, including *not applicable or never* (score=0), *sometimes* (score=1), and *daily* (score=2). There were varied responses for home computer (mean 1.14, SD 0.84), work computer (mean 0.70, SD 0.90), public computer (mean 0.16, SD 0.39), and mobile (mean 1.27, SD 0.85) access.

Health Need

Health needs were operationalized as perceived general health [14]. Thus, it was measured with a single item: “In general, how would you say your health is?” The item was measured on a 5-point scale from *excellent* (score=1) to *poor* (score=5; mean 2.58, SD 0.94). As measured, greater values represent greater health needs or poorer general health.

OHIS Measure

Participants reported using a single item, whether they used a computer, smartphone, or other electronic means to look for health or medical information for themselves in the past 12 months. Responses were *no* (score=0) or *yes* (score=1; mean 0.71, SD 0.45).

Statistical Analysis

The initial demographic data were cleaned and analyzed using SPSS Statistics (version 27, IBM Corporation; [Multimedia Appendix 1](#)). Three group structural equation modeling models were specified based on the Black, Latine, and White

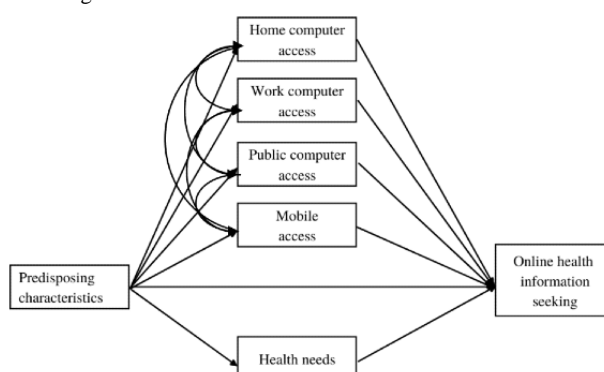
self-categorization using Mplus 8.4 (Muthen and Muthén) [55]. Owing to the dichotomous outcome, diagonally weighted least squares mean and variance adjusted estimators were used instead of maximum likelihood to estimate the models, and odds ratios (ORs; vs standardized coefficients) were used to interpret relationships with OHIS. These models evaluated relationships between predisposing characteristics, access, health needs, and OHIS (H1-H5) and determined the access pairings most predictive of OHIS (RQ1). To test differences across racial and ethnic groups (RQ2), we constrained individual paths and compared each model with its respective baseline model, using chi-square tests for difference testing to account for the diagonally weighted least squares mean and variance estimation method.

Results

Model Fit

Our proposed models grouped by Black, Latin, and White self-categorization displayed poor fit statistics [56]. Modification indices suggested the addition of correlations between all the access variables. Individuals who partake in OHIS are likely to do so in multiple ways [23,57]—thus, these correlations were incorporated into the models ([Figure 2](#)). The resulting models yielded appropriate fit statistics [56] for all 3 models, grouped by Black (root mean square error of approximation [RMSEA]=0.026; comparative fit index [CFI]=0.997; standardized root mean square residual [SRMR]=0.008), Latine (RMSEA=0.021; CFI=0.998; SRMR=0.007), and White (RMSEA=0.026; CFI=0.997; SRMR=0.007) self-categorization.

Figure 2. Final online health information seeking model.



Theoretical Model

First, we examined whether predisposing characteristics were associated with OHIS (H1). Age was negatively associated with OHIS, and education was positively associated with OHIS across all models and groups. Income was positively associated with OHIS for individuals who self-categorized as White (OR 1.122, 95% CI 1.048-1.202; $P<.001$), non-White (OR 1.121, 95% CI 0.989-1.271; $P=.02$), non-Black (OR 1.124, 95% CI 1.052-1.201; $P<.001$), and non-Latine (OR 1.123, 95% CI 1.051-1.200; $P<.001$). Sex was negatively associated with OHIS, such that White (OR 0.572, 95% CI 0.513-0.638; $P<.001$), non-Black (OR 0.591, 95% CI 0.530-0.660; $P<.001$), Latine (OR 0.601, 95% CI 0.474-0.762; $P=.009$), and non-Latine (OR

0.656, 95% CI 0.581-0.741; $P<.001$) females were more likely to engage in OHIS. These findings lent partial support for H1.

Next, we investigated whether predisposing characteristics were associated with access (H2) and whether access was associated with OHIS (H3). Age was negatively associated with all forms of access for all the models and groups. Education was positively associated with all forms of access for all models and groups, except public computer access for Latine individuals ($\beta=.087$; $P=.10$). Income was positively associated with home computer access, work computer access, and mobile access—but negatively associated with public computer access—for all models and groups. Sex was negatively associated with mobile access for all models and groups and negatively associated with work computer access for Black individuals ($\beta=-.097$; $P=.003$),

such that females who self-categorized with these groups accessed the internet more frequently than males within these context-device pairings. However, sex was positively associated with home computer access for individuals who self-categorized as White ($\beta=.071$; $P<.001$), non-Black ($\beta=.074$; $P<.001$), and non-Latine ($\beta=.063$; $P<.001$) as well as public computer access for individuals who self-categorized as non-Black ($\beta=.035$; $P=.04$) and non-Latine ($\beta=.033$; $P=.048$), such that males who self-categorized with these groups accessed the internet more frequently than females. This provided partial support for H2.

Turning to OHIS, mobile and home access were positively associated with OHIS across all models and groups. Public computer access was positively associated with OHIS for non-White (OR 1.307, 95% CI 0.706-2.418; $P=.04$) individuals. Finally, work computer access was not associated with OHIS in any model and for any group. This provided partial support for H3. Regarding RQ1, mobile and home computer access were associated with OHIS more consistently across groups than public computer access and work computer access.

Then, we examined whether predisposing characteristics were associated with health needs (H4) and whether health needs were associated with OHIS (H5). Age was positively associated with health needs across all models and groups. Education and income were negatively associated with health needs across all models and groups, except for the relationship between education and health needs for non-White ($\beta=-.061$; $P=.08$) and Black ($\beta=-.072$; $P=.06$) individuals. Sex was not associated with health needs for any group. This provided partial support for H4. Turning to OHIS, health needs were positively associated with OHIS across all models and groups apart from non-White (OR 1.182, 95% CI 0.958-1.458; $P=.07$) and Latine (OR 1.091, 95% CI 0.869-1.369; $P=.58$) individuals. This provided partial support for H5.

Intersectional Differences

Finally, we investigated whether significant differences in H1 to H5 emerged for different groups (RQ2). We found that certain

relationships between predisposing characteristics and access differed for each type of access; all reported relationships were significant at $P<.05$. Significant differences emerged for the relationship between home computer access and age, income, and sex. The negative association with age was stronger for non-White, Black, and Latine individuals. The positive association with income was stronger for non-White and Black individuals. The association with sex was stronger for non-Black individuals; males were more likely to have home computer access than females among non-Black individuals. Significant differences also emerged for the association of work computer access with age and sex. The negative association with age was significantly stronger for non-Latine individuals. The association with sex was stronger for Black individuals; females were more likely to have work computer access than males among Black individuals. Next, the negative association between public computer access and age was stronger for non-White and Black individuals. Finally, the positive association between mobile access and education was significantly stronger for non-White and Latine individuals.

Other relationships also differed across racial and ethnic groups. For predisposing characteristics and OHIS, sex had a stronger negative association with OHIS for non-Black and White individuals, such that the gap between females and males engaging in OHIS was greater for these groups. For access and OHIS, home computer access had a significantly stronger positive association with OHIS for White, non-Black, and non-Latine individuals. Mobile access had a significantly stronger positive association with OHIS for non-White individuals. There were no significant differences in other dimensions of access or health needs. [Tables 5-7](#) display the ORs and standardized coefficients for the final models, and our general analysis scripts are available in the [Multimedia Appendix 1](#).

Table 5. Standardized coefficients and odds ratios for theorized OHIS^a models (for Black and non-Black individuals)^b.

Path	Group		P value ^e	P value
	Black	Non-Black		
	Standardized coefficient ^c or odds ratio ^d (95% CI)	Standardized coefficient or odds ratio (95% CI)		
Predisposing characteristics → OHIS				
Age (years)	0.982 ^f (0.967-0.996)	0.976 ^f (0.970 -0.982)	.003	<.001
Sex	1.134 ^f (0.721-1.783)	0.591 ^g (0.530-0.660)	.74	<.001
Income	1.081 ^f (0.939-1.245)	1.124 ^f (1.052-1.201)	.15	<.001
Education	1.499 ^f (1.126-1.995)	1.439 ^f (1.259-1.644)	<.001	<.001
Predisposing characteristics → access				
Home computer				
Age (years)	-0.146 ^f	-0.056 ^g	<.001	<.001
Sex	-0.036 ^f	0.074 ^g	.44	<.001
Income	0.329 ^f	0.170 ^g	<.001	<.001
Education	0.213 ^f	0.292 ^f	<.001	<.001
Work computer				
Age (years)	-0.294 ^f	-0.368 ^f	<.001	<.001
Sex	-0.097 ^f	0.018 ^g	.003	.20
Income	0.374 ^f	0.283 ^f	<.001	<.001
Education	0.177 ^f	0.186 ^f	<.001	<.001
Public computer				
Age (years)	-0.255 ^f	-0.180 ^g	<.001	<.001
Sex	0.033 ^f	0.035 ^f	.39	.04
Income	-0.122 ^f	-0.115 ^f	.003	<.001
Education	0.104 ^f	0.135 ^f	.01	<.001
Mobile				
Age (years)	-0.446 ^f	-0.444 ^f	<.001	<.001
Sex	-0.105 ^f	-0.060 ^f	<.001	<.001
Income	0.226 ^f	0.220 ^f	<.001	<.001
Education	0.167 ^f	0.131 ^f	<.001	<.001
Access → OHIS				
Home computer	1.491 ^f (0.990-2.246)	1.981 ^g (1.554-2.526)	.001	<.001
Work computer	0.877 ^f (0.663-1.161)	0.939 ^f (0.822-1.073)	.55	.92
Public computer	1.368 ^f (0.692-2.706)	1.118 ^f (0.801-1.560)	.06	.16
Mobile	2.158 ^f (1.175-3.962)	1.833 ^f (1.440-2.333)	<.001	<.001
Predisposing Characteristics → health need				
Age (years)	0.098 ^f	0.102 ^f	.005	<.001
Sex	0.001 ^f	0.022 ^f	.98	.16

Path	Group			
	Black	Non-Black		
	Standardized coefficient ^c or odds ratio ^d (95% CI)	<i>P</i> value ^e	Standardized coefficient or odds ratio (95% CI)	<i>P</i> value
Income	-0.202 ^f	<.001	-0.245 ^f	<.001
Education	-0.072 ^f	.06	-0.117 ^f	<.001
Health need → OHIS	1.235 ^f (0.970-1.572)	.03	1.222 ^f (1.078-1.385)	.004

^aOHIS: online health information seeking.

^bComparisons were made for each model per row.

^cStandardized coefficients are displayed for paths predicting nondichotomous outcomes; negative relationships are indicated by negative signs. For standardized coefficients, 95% CI values are not available.

^dOdds ratios are presented for paths predicting dichotomous outcomes (ie, OHIS) and were generated using Monte Carlo integration because of model complexity; negative relationships are indicated by values <1.

^eSignificance values were based on the primary models (ie, without Monte Carlo integration).

^fCoefficients or odds ratios differ significantly from those denoted by footnote *g* at $P < .05$.

^gCoefficients or odds ratios differ significantly from those denoted by footnote *f* at $P < .05$.

Table 6. Standardized coefficients and odds ratios for theorized OHIS^a models (for Latine and non-Latine individuals)^b.

Path	Group		P value ^e	P value
	Latine	Non-Latine		
	Standardized coefficient ^c or odds ratio ^d (95% CI)	Standardized coefficient or odds ratio (95% CI)		
Predisposing characteristics → OHIS				
Age (years)	0.979 ^f (0.966-0.993)	0.977 ^f (0.971-0.983)	.002	<.001
Sex	0.601 ^f (0.474-0.762)	0.656 ^f (0.581-0.741)	.009	<.001
Income	1.058 ^f (0.930-1.204)	1.123 ^f (1.051-1.200)	.42	<.001
Education	1.393 ^f (1.075-1.804)	1.481 ^f (1.291-1.699)	<.001	<.001
Predisposing characteristics → access				
Home computer				
Age (years)	-0.185 ^f	-0.054 ^g	<.001	<.001
Sex	0.060 ^f	0.063 ^f	.08	<.001
Income	0.181 ^f	0.203 ^f	<.001	<.001
Education	0.324 ^f	0.254 ^f	<.001	<.001
Work computer				
Age (years)	-0.249 ^f	-0.374 ^g	<.001	<.001
Sex	0.019 ^f	-0.006 ^f	.57	.68
Income	0.265 ^f	0.301 ^f	<.001	<.001
Education	0.259 ^f	0.171 ^f	<.001	<.001
Public computer				
Age (years)	-0.257 ^f	-0.188 ^f	<.001	<.001
Sex	-0.024 ^f	0.033 ^f	.57	.048
Income	-0.152 ^f	-0.142 ^f	.002	<.001
Education	0.087 ^f	0.128 ^f	.10	<.001
Mobile				
Age (years)	-0.462 ^f	-0.439 ^f	<.001	<.001
Sex	-0.069 ^f	-0.072 ^f	.03	<.001
Income	0.139 ^f	0.232 ^f	<.001	<.001
Education	0.206 ^f	0.114 ^g	<.001	<.001
Access → OHIS				
Home computer	1.294 ^f (0.861-1.945)	1.950 ^g (1.541-2.467)	.04	<.001
Work computer	1.121 ^f (0.772-1.627)	0.917 ^f (0.806-1.044)	.19	.49
Public computer	1.500 ^f (0.574-3.919)	1.117 ^f (0.824-1.513)	.08	.15
Mobile	1.739 ^f (1.057-2.861)	1.912 ^f (1.491-2.452)	<.001	<.001
Predisposing characteristics → health need				
Age (years)	0.140 ^f	0.093 ^f	<.001	<.001
Sex	-0.039 ^f	0.028 ^f	.30	.07

Path	Group		P value ^e	P value
	Latine	Non-Latine		
	Standardized coefficient ^c or odds ratio ^d (95% CI)	Standardized coefficient or odds ratio (95% CI)		
Income	-0.178 ^f	-0.243 ^f	<.001	<.001
Education	-0.147 ^f	-0.113 ^f	<.001	<.001
Health need → OHIS	1.091 ^f (0.869-1.369)	1.242 ^f (1.093-1.411)	.58	.001

^aOHIS: online health information seeking.

^bComparisons were made for each model per row.

^cStandardized coefficients are displayed for paths predicting nondichotomous outcomes; negative relationships are indicated by negative signs. For standardized coefficients, 95% CI values are not available.

^dOdds ratios are presented for paths predicting dichotomous outcomes (ie, OHIS) and were generated using Monte Carlo integration because of model complexity; negative relationships are indicated by values <1.

^eSignificance values were based on the primary models (ie, without Monte Carlo integration).

^fCoefficients or odds ratios differ significantly from those denoted by footnote *g* at $P < .05$.

^gCoefficients or odds ratios differ significantly from those denoted by footnote *f* at $P < .05$.

Table 7. Standardized coefficients and odds ratios for theorized OHIS^a models (for White and non-White individuals)^b.

Path	Group		P value ^e	Standardized coefficient or odds ratio (95% CI)	P value
	White	Non-White			
Predisposing characteristics → OHIS					
Age (years)	0.975 ^f (0.967-0.983)		<.001	0.982 ^f (0.971-0.994)	<.001
Sex	0.572 ^f (0.513-0.638)		<.001	1.123 ^g (0.757-1.665)	.84
Income	1.122 ^f (1.048-1.202)		<.001	1.121 ^f (0.989-1.271)	.02
Education	1.456 ^f (1.267-1.673)		<.001	1.427 ^f (1.119-1.819)	<.001
Predisposing characteristics → access					
Home computer					
Age (years)	-0.047 ^f		.003	-0.158 ^g	<.001
Sex	0.071 ^f		<.001	0.014 ^f	.62
Income	0.176 ^f		<.001	0.265 ^g	<.001
Education	0.282 ^f		<.001	0.280 ^f	<.001
Work computer					
Age (years)	-0.370 ^f		<.001	-0.298 ^f	<.001
Sex	-0.001 ^f		.95	0.010 ^f	.71
Income	0.289 ^f		<.001	0.329 ^f	<.001
Education	0.182 ^f		<.001	0.205 ^f	<.001
Public computer					
Age (years)	-0.178 ^f		<.001	-0.242 ^g	<.001
Sex	0.033 ^f		.06	0.038 ^f	.26
Income	-0.122 ^f		<.001	-0.129 ^f	<.001
Education	0.131 ^f		<.001	0.126 ^f	.001
Mobile					
Age (years)	-0.450 ^f		<.001	-0.432 ^f	<.001
Sex	-0.069 ^f		<.001	-0.065 ^f	.01
Income	0.223 ^f		<.001	0.202 ^f	<.001
Education	0.121 ^f		<.001	0.199 ^g	<.001
Access → OHIS					
Home computer	2.002 ^f (1.558-2.573)		<.001	1.458 ^g (1.005-2.116)	<.001
Work computer	0.921 ^f (0.806-1.052)		.63	0.931 ^f (0.712-1.218)	.96
Public computer	1.098 ^f (0.787-1.532)		.25	1.307 ^f (0.706-2.418)	.037
Mobile	1.776 ^f (1.398-2.256)		<.001	2.379 ^g (1.281-4.419)	<.001
Predisposing characteristics → health need					
Age (years)	0.097 ^f		<.001	0.123 ^f	<.001
Sex	0.027 ^f		.09	-0.015 ^f	.64

Path	Group		Standardized coefficient or odds ratio (95% CI)	P value ^e	Standardized coefficient or odds ratio (95% CI)	P value
	White	Non-White				
Income	-0.246 ^f			<.001	-0.226 ^f	<.001
Education	-0.124 ^f			<.001	-0.061 ^f	.08
Health need → OHIS	1.237 ^f (1.085-1.411)			.002	1.182 ^f (0.958-1.458)	.07

^aOHIS: online health information seeking.

^bComparisons were made for each model per row.

^cStandardized coefficients are displayed for paths predicting nondichotomous outcomes; negative relationships are indicated by negative signs. For standardized coefficients, 95% CI values are not available.

^dOdds ratios are presented for paths predicting dichotomous outcomes (ie, OHIS) and were generated using Monte Carlo integration because of model complexity; negative relationships are indicated by values <1.

^eSignificance values were based on the primary models (ie, without Monte Carlo integration).

^fCoefficients or odds ratios differ significantly from those denoted by footnote *g* at $P < .05$.

^gCoefficients or odds ratios differ significantly from those denoted by footnote *f* at $P < .05$.

Discussion

Principal Findings

This study applied a nuanced conceptualization of access to theoretical models of OHIS and identified how relationships with OHIS differed between racial and ethnic groups (ie, Black, Latine, and White individuals). We found partial support for all hypotheses, and results regarding the RQs provided deeper insight into the predicted relationships. By examining access as 4 unique context–device pairings, we found that home computer and mobile access were most consistently associated with OHIS. In addition, disaggregating models by racial and ethnic self-categorization identified different patterns between predisposing characteristics and access for different groups, highlighting how the digital divide affects intersectional groups.

Our findings suggest that predisposing characteristics are associated with OHIS for different racial and ethnic groups (H1). Education was positively associated with OHIS, and age was negatively associated with OHIS. These findings align with previous research, such that those with more education and younger individuals are more likely to possess the skills to navigate web-based platforms [30,31,33,58]. However, income only had a positive association with OHIS for individuals who self-categorized as White, non-Black, or non-Latine. Although education and income are often correlated with OHIS [9,59,60], our findings suggest that education may better index the foundational knowledge necessary to take advantage of web-based health information. Finally, although females sought out health information more frequently than males [23,32], this pattern did not hold for those who self-categorized as Black. This may reflect how factors of socioeconomic status are typically stronger determinants of access to technology and health services for racial minorities than sociodemographic factors [44,61].

Our findings also suggest that some predisposing characteristics are associated with access for some racial and ethnic groups (H2). Age was negatively associated with all forms of access.

Older individuals used all 4 context–device pairings less frequently than younger individuals, which may indicate their use of nondigital means (eg, print media and interpersonal) to obtain health information [47]. Females of all groups accessed the internet on mobile devices more frequently than males, as well as work computers among Black individuals. However, White males (vs females) accessed the internet on home computers more frequently. Although males and females may have similar access overall [62], combining a multifaceted conceptualization of access with an intersectional approach highlights that disparities in access based on sex are grounded in devices and contexts of use, as well as race and ethnicity. White males and females are more likely to access the internet on home computers and mobile devices, respectively, suggesting that internet access may be a zero-sum game, such that having access in one place reduces the need to have access elsewhere [21]. However, this trade-off did not emerge for other groups, suggesting important boundary conditions based on race and ethnicity [61]. Furthermore, income and education consistently demonstrated a positive association with access, as maintaining access requires sustainable resources afforded by income and education [7]. However, income was negatively associated with public computer access, suggesting that individuals with lower income may be more reliant on public resources to access the internet [39,57].

Our findings generally confirm that access is associated with OHIS (H3). As suggested by previous research [14,28,50-52], OHIS is unlikely without a means to access the internet. Specifically, mobile access was positively associated with OHIS for all groups, suggesting that the ubiquity of mobile phones may help bridge this particular gap of the digital divide [3,5,21]. Home computer access was also associated with OHIS for all groups. Public computer access was positively associated with OHIS for non-White individuals. Work computer access was not associated with OHIS across all groups. These findings are corroborated by the fact that certain contexts of OHIS (eg, home computer and mobile) provide a level of privacy that other contexts do not [23,37], thus facilitating searches for private

health information. Our results highlight that certain groups (particularly non-White individuals) face access disparities based on affordance and maintenance of that privacy [7,63].

Predisposing characteristics were also associated with health needs (H4), such that older individuals and individuals with less education and income were more likely to describe their health as poor. Older individuals and individuals with less education and income often face barriers to quality health options [44]. However, education was not significantly associated with health needs for Black and non-White individuals. Research on minority groups (eg, Black individuals) finds that educational advancement may not overcome the aggregated stress of marginalization, which contributes to negative health outcomes [64]. However, sex was not associated with health needs for any group, likely because of counterbalancing of health issues that disproportionately affect males and females separately [45].

Furthermore, those with greater health needs were more likely to partake in OHIS, apart from non-White and Latine individuals (H5). Past research has found that greater health needs are associated with increased OHIS among Latine individuals [58]. However, our findings support previous findings that Latine individuals may be less trusting of health information on the web and may rely on different (eg, interpersonal) means of seeking out health information [65]. Reconciliation of these contradicting findings may be a result of area, as the national sample is not limited to patterns that may only exist in larger cities with more resources to provide access [58].

Finally, our exploratory analyses provide insight into RQ2; however, additional research may be required to fully explicate certain patterns in our model in which stronger relationships were detected for specific racial and ethnic groups. In terms of access, several relationships were stronger for Black individuals. Greater income was associated with more frequent home computer access across all groups; however, this relationship was stronger for Black (vs non-Black) individuals. Income inequality among Black individuals appears to be a stark determinant of home computer access [39]. Individuals with higher income can afford the cost of maintenance that comes with home computer access, which is apparent across the models [7]. However, Black individuals with lower income may face additional hurdles to home computer access, such as living in areas without the infrastructure to support maintenance [3]. Similarly, the negative relationship between age and public computer access was stronger for Black (vs non-Black) individuals. The restricted availability of public computers [5] and limited accessibility of web-based platforms for older individuals [7] may be particularly profound for Black individuals. As Black individuals have been historically disadvantaged in access, both community resources and technological skills of the older generation may be stunted [5,39]. Furthermore, the relationship between sex and access differed, such that Black females reported more frequent access via work computers than Black males. In contrast, non-Black males were more likely to use home computers. For specific sex, racial, and ethnic groups, finding access to the internet via 1 mode may be sufficient, which could reduce the need to have access elsewhere [21]. Finally, non-White (vs White)

individuals, or racial and ethnic minorities in general, demonstrated a stronger negative relationship between age and home computer access and a stronger positive relationship between education and mobile access. Older racial and ethnic minorities tend to have less access to the internet [39], including at home. Although racial and ethnic minorities lag in home computer ownership, the stronger relationship with education and mobile access may be interpreted as a route to attenuate the digital divide or as exacerbating the digital divide within racial minority groups. As such, lower education levels seem to inhibit access more intensely among Latine individuals.

In addition to access, the relationship between sex and OHIS differed, such that non-Black (vs Black) females demonstrated a stronger association with OHIS. The extent to which females relieve the burden of family health knowledge [32] may differ across racial and ethnic groups, as these groups are often disproportionately affected by health disparities [4,6]. Moreover, the positive relationship between home computer access and OHIS was weaker for non-White (vs White) individuals, whereas the positive relationship between mobile access and OHIS was stronger. Mobile devices remained a key factor not only for establishing access for racial and ethnic minorities [3] but also for OHIS. Although the mobile interface may be more difficult for tasks such as OHIS [8], having at least one point of access is critical for web-based health behaviors [9,24]. Although some OHIS research suggests that access has been democratized [19], the above results highlight the overlap in health and technology disparities for racial and ethnic minority groups [4-6].

Contributions

Our first contribution—applying a multidimensional conceptualization of access to theoretical models of OHIS—revealed that different context-device pairings offer distinct OHIS profiles. Mobile and home computer access were more consistently associated with OHIS than work computer and public computer access. This implies that privacy is important when assessing the digital divide, as home computers and mobile devices can be used in more private contexts [23,37]. Furthermore, for racial and ethnic minority participants, the link between home computer access and OHIS was weaker, and the relationship between mobile access and OHIS was stronger. These differences primarily emerged in the non-White versus White models because of reduced power in the other models and the fact that Black individuals were included as non-Latine individuals (and vice versa). Racial and ethnic minorities may rely more on mobile (vs home computer) access for OHIS because of the lower cost and flexibility of mobile devices [66]. Public and work computer access were not consistently associated with OHIS, although public computer access was generally associated with OHIS for non-White individuals. Contexts that typically do not require ownership may provide access for groups that lack other means of access. Discrepancies among home computer, work computer, public computer, and mobile access highlight that devices and contexts of use do not provide access equally [7].

Our second contribution was to unpack the digital divide using an intersectional approach, as it is crucial to understand which

groups have limited access to the internet. We found discrepancies in access for specific groups. Older individuals who self-categorized as a racial or ethnic minority engaged in less frequent home and public computer access. Older (vs younger) individuals and racial and ethnic minorities (vs majorities) tend to access the internet less frequently [39], and we find that this gap is magnified for home and public computer access. In addition, for non-White (vs White) individuals, there were stronger positive relationships between education and mobile access, as well as mobile access and OHIS. Although formal education may minimize the digital divide via mobile access, disparities in access to the education needed to operate the devices should also be considered. Discrepancies between predisposing characteristics, access, health needs, and OHIS for different racial and ethnic groups demonstrate the need for future OHIS theorizing to adopt an intersectional approach.

Practical Implications

This study supports the criticism that the digital divide is not a dichotomy between access and lack thereof [22]. In response to this criticism, interventions combating the digital divide must consider how context–device access pairings can be leveraged among specific racial and ethnic groups. Home computer and mobile were the most frequently used means of access and were both consistently positively related to OHIS. This implies that people typically engage in OHIS on home computers or mobile devices. As such, improving access within these contexts may be valuable for interventions to support OHIS across racial and ethnic groups. Our findings suggest that work and public computer access are less ideal for OHIS. These pairings may lack accessibility [5] or privacy [23,37]. The association between public computer access and OHIS for racial and ethnic minorities could be explored further as a means of increasing OHIS for Latine and Black individuals. Work and public computer access remain important to the extent that increasing access points overall supports OHIS [21,25,26]. However, public computer access may also replace more expensive yet more private modes of access (eg, home computer). In general, interventions could reinforce existing strengths (eg, home computer and mobile access) or bolster existing weaknesses (eg, work and public computer access).

These 2 courses of action can also apply to future interventions aimed at addressing the digital divide and OHIS among specific groups. For racial and ethnic minorities, we found weaker positive relationships between home computer access and OHIS and stronger positive relationships between mobile access and OHIS. Interventions can strengthen the established relationship between mobile devices and OHIS or bolster the weaker link for home computer access. Although home computer access is considered a more easily navigable interface [8], it may not be scalable, given its cost. As such, interventions with limited financial resources may benefit from working with providers of web-based health information to develop mobile-friendly interfaces to make health information more accessible.

Furthermore, this study shines a spotlight on older racial and ethnic minorities, who experienced consistent discrepancies in access and may have the highest health needs. The negative relationship between age and public computer access was stronger for Black individuals, suggesting a drop-off in public computer access for older Black adults. As public computer access was associated with OHIS for racial and ethnic minorities, future interventions could increase the accessibility of public computers for older Black adults, with attention toward local libraries and community centers in predominately older Black neighborhoods [5]. Overall, a deep understanding of current community strengths must be balanced with efforts to provide equitable access to web-based health information to not overlook the key interactions between multiple social positions that create compounding experiences of oppression [67].

Limitations and Future Research

Some limitations should be considered when interpreting these findings. This study used secondary cross-sectional data. Thus, potentially relevant variables (eg, mobile use at home vs at work vs in public) were not measured, and causality or directionality cannot be determined. Future research could measure additional constructs and use longitudinal designs. In addition, this analysis used self-reported data. Future work could use log and GPS data in tandem to paint a more accurate picture of OHIS. Furthermore, our primary outcome variable (OHIS) was dichotomous, and other variables (eg, access) were trichotomous or single-item measures. Future research should use continuous variables for OHIS and access to better capture the temporal variety of digital media use. Finally, we did not examine second-level digital divide variables (eg, experience, perceived utility, beliefs, and skills) [28,50,51,68,69]. However, before receiving and interpreting information, those who seek information on the web must choose a device and context to seek that information.

Conclusions

This study holds that a nuanced conceptualization of access is necessary to understand how the digital divide differentially affects racial and ethnic groups. Our theoretical model identified variables that predict OHIS while distinguishing the type (ie, device) and location (ie, context) of access, testing these associations for different racial and ethnic groups and examining intersectional characteristics among these groups (ie, age, sex, education, and income). By interlacing a thorough understanding of the first-level digital divide with an awareness of the unique impacts of the digital divide for specific groups, we further theorize on OHIS and suggest important considerations for more targeted interventions. As we continue to understand the complexities of the digital divide and its relationship with health, racial, and ethnic disparities, our perspective highlights how web-based health resources may not be accessed by those who need them the most.

Acknowledgments

The authors would like to thank Dr. David DeAndrea and Dr. Kelly Garrett for assisting them in the publication process, and Dr. Kim for teaching them structural equation modeling (SEM) and giving them the space to grow on their own.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Data and syntax.

[[ZIP File \(Zip Archive\), 227 KB - jmir_v24i3e32678_app1.zip](#)]

References

1. Eysenbach G. What is e-health? *J Med Internet Res* 2001 Jun;3(2):E20 [FREE Full text] [doi: [10.2196/jmir.3.2.e20](https://doi.org/10.2196/jmir.3.2.e20)] [Medline: [11720962](https://pubmed.ncbi.nlm.nih.gov/11720962/)]
2. Falling through the net: defining the digital divide. National Telecommunications and Information Administration. 1999. URL: <https://ictlogy.net/bibliography/reports/projects.php?idp=590> [accessed 2021-03-23]
3. Smartphones help blacks, Hispanics bridge some – but not all – digital gaps with whites. Pew Research Center. 2019 Aug 20. URL: <https://www.benton.org/headlines/smartphones-help-blacks-hispanics-bridge-some-%E2%80%93-not-all-%E2%80%93-digital-gaps-whites> [accessed 2021-02-10]
4. Massey DS, Denton NA. *American Apartheid – Segregation and the Making of the Underclass*. Cambridge, MA: Harvard University Press; 1993.
5. Mossberger K, Tolbert CJ, Gilbert M. Race, place, and information technology. *Urban Affairs Rev* 2016 Aug 19;41(5):583-620. [doi: [10.1177/1078087405283511](https://doi.org/10.1177/1078087405283511)]
6. Sewell S, Massey D, Denton N. A different menu: racial residential segregation the persistence of racial inequality. In: Ray R, editor. *Race and Ethnic Relations in the 21st Century: History, Theory, Institutions, and Policy*. San Diego, CA: University Readers; 2010:287-296.
7. van Deursen AJ, van Dijk JA. The first-level digital divide shifts from inequalities in physical access to inequalities in material access. *New Media Soc* 2019 Feb 07;21(2):354-375 [FREE Full text] [doi: [10.1177/1461444818797082](https://doi.org/10.1177/1461444818797082)] [Medline: [30886536](https://pubmed.ncbi.nlm.nih.gov/30886536/)]
8. Napoli PM, Obar JA. The emerging mobile internet underclass: a critique of mobile internet access. *Inform Society* 2014 Oct 02;30(5):323-334. [doi: [10.1080/01972243.2014.944726](https://doi.org/10.1080/01972243.2014.944726)]
9. Gonzales AL, Ems L, Suri VR. Cell phone disconnection disrupts access to healthcare and health resources: a technology maintenance perspective. *New Media Society* 2016 Jul 10;18(8):1422-1438. [doi: [10.1177/1461444814558670](https://doi.org/10.1177/1461444814558670)]
10. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. *Health Mark Q* 2008 Jul;25(1-2):175-203. [doi: [10.1080/07359680802126301](https://doi.org/10.1080/07359680802126301)] [Medline: [18935884](https://pubmed.ncbi.nlm.nih.gov/18935884/)]
11. Public use dataset: HINTS data terms of use. Health Information National Trends Survey (HINTS). 2019. URL: <https://hints.cancer.gov/data/download-data.aspx> [accessed 2020-09-19]
12. Fox S, Duggan M. Health online 2013. Pew Research Center. 2013. URL: <http://www.pewinternet.org/2013/01/15/health-online-2013/> [accessed 2020-11-13]
13. Lambert SD, Loisel CG. Health information seeking behavior. *Qual Health Res* 2007 Oct;17(8):1006-1019. [doi: [10.1177/1049732307305199](https://doi.org/10.1177/1049732307305199)] [Medline: [17928475](https://pubmed.ncbi.nlm.nih.gov/17928475/)]
14. Chisolm DJ. Does online health information seeking act like a health behavior?: a test of the behavioral model. *Telemed J E Health* 2010 Mar;16(2):154-160. [doi: [10.1089/tmj.2009.0102](https://doi.org/10.1089/tmj.2009.0102)] [Medline: [20156127](https://pubmed.ncbi.nlm.nih.gov/20156127/)]
15. Gonzales A. The contemporary US digital divide: from initial access to technology maintenance. *Inform Commun Society* 2015 Jun 02;19(2):234-248. [doi: [10.1080/1369118x.2015.1050438](https://doi.org/10.1080/1369118x.2015.1050438)]
16. Masi CM, Suarez-Balcazar Y, Cassey MZ, Kinney L, Piotrowski ZH. Internet access and empowerment: a community-based health initiative. *J Gen Intern Med* 2003 Jul;18(7):525-530 [FREE Full text] [doi: [10.1046/j.1525-1497.2003.20344.x](https://doi.org/10.1046/j.1525-1497.2003.20344.x)] [Medline: [12848835](https://pubmed.ncbi.nlm.nih.gov/12848835/)]
17. Barzilai-Nahon K. Gaps and bits: conceptualizing measurements for digital divide/s. *Inform Society* 2006 Dec;22(5):269-278. [doi: [10.1080/01972240600903953](https://doi.org/10.1080/01972240600903953)]
18. van Deursen AJ, van Dijk JA. Toward a multifaceted model of internet access for understanding digital divides: an empirical investigation. *Inform Society* 2015 Sep 11;31(5):379-391. [doi: [10.1080/01972243.2015.1069770](https://doi.org/10.1080/01972243.2015.1069770)]
19. Ducrot P, Montagni I, Thanh VN, Serry A, Richard J. Evolution of online health-related information seeking in France from 2010 to 2017: results from nationally representative surveys. *J Med Internet Res* 2021 Apr 14;23(4):e18799 [FREE Full text] [doi: [10.2196/18799](https://doi.org/10.2196/18799)] [Medline: [33851927](https://pubmed.ncbi.nlm.nih.gov/33851927/)]

20. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The digital divide in health-related technology use: the significance of race/ethnicity. *Gerontologist* 2019 Jan 09;59(1):6-14. [doi: [10.1093/geront/gny138](https://doi.org/10.1093/geront/gny138)] [Medline: [30452660](https://pubmed.ncbi.nlm.nih.gov/30452660/)]
21. Reisdorf BC, Fernandez L, Hampton KN, Shin I, Dutton WH. Mobile phones will not eliminate digital and social divides: how variation in internet activities mediates the relationship between type of internet access and local social capital in Detroit. *Soc Sci Comput Rev* 2020 Mar 16;089443932090944. [doi: [10.1177/0894439320909446](https://doi.org/10.1177/0894439320909446)]
22. Gunkel DJ. Second thoughts: toward a critique of the digital divide. *New Media Society* 2016 Jun 30;5(4):499-522. [doi: [10.1177/146144480354003](https://doi.org/10.1177/146144480354003)]
23. Hassani SN. Locating digital divides at home, work, and everywhere else. *Poetics* 2006 Aug;34(4-5):250-272. [doi: [10.1016/j.poetic.2006.05.007](https://doi.org/10.1016/j.poetic.2006.05.007)]
24. Mossberger K, Tolbert C, Hamilton A. Broadband adoption measuring digital citizenship: mobile access and broadband. *Int J Commun* 2012;2492-2528 [FREE Full text]
25. Dobransky K, Hargittai E. Inquiring minds acquiring wellness: uses of online and offline sources for health information. *Health Commun* 2012;27(4):331-343. [doi: [10.1080/10410236.2011.585451](https://doi.org/10.1080/10410236.2011.585451)] [Medline: [21932982](https://pubmed.ncbi.nlm.nih.gov/21932982/)]
26. Percheski C, Hargittai E. Health information-seeking in the digital age. *J Am Coll Health* 2011 Apr 13;59(5):379-386. [doi: [10.1080/07448481.2010.513406](https://doi.org/10.1080/07448481.2010.513406)] [Medline: [21500056](https://pubmed.ncbi.nlm.nih.gov/21500056/)]
27. Alper M, Katz VS, Clark LS. Researching children, intersectionality, and diversity in the digital age. *J Child Media* 2016 Jan 18;10(1):107-114. [doi: [10.1080/17482798.2015.1121886](https://doi.org/10.1080/17482798.2015.1121886)]
28. Marton C. Understanding how women seek health information on the web. Doctoral Dissertation - University of Toronto. 2011. URL: <https://tspace.library.utoronto.ca/handle/1807/29808> [accessed 2022-02-09]
29. Cotten SR, Gupta SS. Characteristics of online and offline health information seekers and factors that discriminate between them. *Soc Sci Med* 2004 Nov;59(9):1795-1806. [doi: [10.1016/j.socscimed.2004.02.020](https://doi.org/10.1016/j.socscimed.2004.02.020)] [Medline: [15312915](https://pubmed.ncbi.nlm.nih.gov/15312915/)]
30. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. *J Med Internet Res* 2017 Jun 19;19(6):e218 [FREE Full text] [doi: [10.2196/jmir.7579](https://doi.org/10.2196/jmir.7579)] [Medline: [28630033](https://pubmed.ncbi.nlm.nih.gov/28630033/)]
31. Wong DK, Cheung M. Online health information seeking and eHealth literacy among patients attending a primary care clinic in Hong Kong: a cross-sectional survey. *J Med Internet Res* 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: [10.2196/10831](https://doi.org/10.2196/10831)] [Medline: [30916666](https://pubmed.ncbi.nlm.nih.gov/30916666/)]
32. 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](https://doi.org/10.2196/jmir.6668)] [Medline: [28377367](https://pubmed.ncbi.nlm.nih.gov/28377367/)]
33. Bjarnadottir RI, Millery M, Fleck E, Bakken S. Correlates of online health information-seeking behaviors in a low-income Hispanic community. *Inform Health Soc Care* 2016 Dec;41(4):341-349. [doi: [10.3109/17538157.2015.1064429](https://doi.org/10.3109/17538157.2015.1064429)] [Medline: [26837012](https://pubmed.ncbi.nlm.nih.gov/26837012/)]
34. Li J, Theng Y, Foo S. Predictors of online health information seeking behavior: changes between 2002 and 2012. *Health Informatics J* 2016;22(4):804-814. [doi: [10.1177/1460458215595851](https://doi.org/10.1177/1460458215595851)] [Medline: [26261218](https://pubmed.ncbi.nlm.nih.gov/26261218/)]
35. Humphreys L, Karnowski V, von Thilo P. Smartphones as metamedia: a framework for identifying the niches structuring smartphone use. *Int J Commun* 2018;12(17):2793-2809 [FREE Full text]
36. Kaplan D, Mossberger K. Prospects for poor neighborhoods in the broadband era. *Econ Dev Q* 2011 Dec 23;26(1):95-105. [doi: [10.1177/0891242411431450](https://doi.org/10.1177/0891242411431450)]
37. Collins SA, Yoon S, Rockoff ML, Nocenti D, Bakken S. Digital divide and information needs for improving family support among the poor and underserved. *Health Informatics J* 2016 Mar 15;22(1):67-77 [FREE Full text] [doi: [10.1177/1460458214536065](https://doi.org/10.1177/1460458214536065)] [Medline: [24935213](https://pubmed.ncbi.nlm.nih.gov/24935213/)]
38. Campbell SW. Mobile media and communication: a new field, or just a new journal? *Mobile Media Commun* 2013 Jan 01;1(1):8-13. [doi: [10.1177/2050157912459495](https://doi.org/10.1177/2050157912459495)]
39. Smith A. US smartphone use in 2015. Pew Research Center. 2015. URL: <https://www.pewresearch.org/internet/2015/04/01/us-smartphone-use-in-2015/> [accessed 2021-01-15]
40. Cline R, Haynes K. 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](https://doi.org/10.1093/her/16.6.671)] [Medline: [11780707](https://pubmed.ncbi.nlm.nih.gov/11780707/)]
41. Fareed N, Swoboda CM, Jonnalagadda P, Walker DM, Huerta TR. Differences between races in health information seeking and trust over time: evidence from a cross-sectional, pooled analyses of HINTS data. *Am J Health Promot* 2021 Jan 26;35(1):84-92. [doi: [10.1177/0890117120934609](https://doi.org/10.1177/0890117120934609)] [Medline: [32588638](https://pubmed.ncbi.nlm.nih.gov/32588638/)]
42. Gonzales AL. Health benefits and barriers to cell phone use in low-income urban U.S. neighborhoods: indications of technology maintenance. *Mobile Media Commun* 2014 Aug 19;2(3):233-248. [doi: [10.1177/2050157914530297](https://doi.org/10.1177/2050157914530297)]
43. van Dijk JA. *The Deepening Divide: Inequality in the Information Society*. Thousand Oaks, CA: Sage Publications; 2005.
44. Braveman P, Egerter S, Williams DR. The social determinants of health: coming of age. *Annu Rev Public Health* 2011 Apr 21;32(1):381-398. [doi: [10.1146/annurev-publhealth-031210-101218](https://doi.org/10.1146/annurev-publhealth-031210-101218)] [Medline: [21091195](https://pubmed.ncbi.nlm.nih.gov/21091195/)]
45. Chapman EN, Kaatz A, Carnes M. Physicians and implicit bias: how doctors may unwittingly perpetuate health care disparities. *J Gen Intern Med* 2013 Nov 11;28(11):1504-1510 [FREE Full text] [doi: [10.1007/s11606-013-2441-1](https://doi.org/10.1007/s11606-013-2441-1)] [Medline: [23576243](https://pubmed.ncbi.nlm.nih.gov/23576243/)]

46. Sood R, Jenkins SM, Sood A, Clark MM. Gender differences in self-perception of health at a wellness center. *Am J Health Behav* 2019 Nov 01;43(6):1129-1135. [doi: [10.5993/ajhb.43.6.10](https://doi.org/10.5993/ajhb.43.6.10)]
47. Fang M, Canham S, Battersby L, Sixsmith J, Wada M, Sixsmith A. Exploring privilege in the digital divide: implications for theory, policy, and practice. *Gerontologist* 2019 Jan 09;59(1):1-15. [doi: [10.1093/geront/gny037](https://doi.org/10.1093/geront/gny037)] [Medline: [29750241](https://pubmed.ncbi.nlm.nih.gov/29750241/)]
48. LaVeist TA, Gaskin D, Richard P. Estimating the economic burden of racial health inequalities in the United States. *Int J Health Serv* 2011 Apr 01;41(2):231-238. [doi: [10.2190/hs.41.2.c](https://doi.org/10.2190/hs.41.2.c)]
49. Crenshaw K. Demarginalizing the intersection of race and sex: a Black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. In: Levit N, Verchick RR, editors. *Feminist Legal Theory*. New York, NY: Routledge; 1989.
50. Reifegerste D, Blech S, Dechant P. Understanding information seeking about the health of others: applying the comprehensive model of information seeking to proxy online health information seeking. *J Health Commun* 2020 Feb 01;25(2):126-135. [doi: [10.1080/10810730.2020.1716280](https://doi.org/10.1080/10810730.2020.1716280)] [Medline: [32009552](https://pubmed.ncbi.nlm.nih.gov/32009552/)]
51. van Stee SK, Yang Q. Online cancer information seeking: applying and extending the comprehensive model of information seeking. *Health Commun* 2018 Dec 30;33(12):1583-1592. [doi: [10.1080/10410236.2017.1384350](https://doi.org/10.1080/10410236.2017.1384350)] [Medline: [29083231](https://pubmed.ncbi.nlm.nih.gov/29083231/)]
52. Basnyat I, Nekmat E, Jiang S, Lin J. Applying the modified comprehensive model of information seeking to online health information seeking in the context of India. *J Health Commun* 2018 Jul 06;23(6):563-572. [doi: [10.1080/10810730.2018.1493058](https://doi.org/10.1080/10810730.2018.1493058)] [Medline: [29979921](https://pubmed.ncbi.nlm.nih.gov/29979921/)]
53. Health Information National Trends Survey 5 (HINTS 5) Cycle 3 Methodology Report. HINTS. 2019. URL: https://hints.cancer.gov/docs/Instruments/HINTS5_Cycle3_MethodologyReport.pdf [accessed 2020-09-19]
54. Finney Rutten LJ, Blake KD, Skolnick VG, Davis T, Moser RP, Hesse BW. Data Resource Profile: The National Cancer Institute's Health Information National Trends Survey (HINTS). *Int J Epidemiol* 2020 Feb 01;49(1):17-17j [FREE Full text] [doi: [10.1093/ije/dyz083](https://doi.org/10.1093/ije/dyz083)] [Medline: [31038687](https://pubmed.ncbi.nlm.nih.gov/31038687/)]
55. Muthén LK. *Mplus: The Comprehensive Modeling Program for Applied Researchers : User's Guide*. Los Angeles, CA: Muthén & Muthén; 1998.
56. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equat Model Multidiscipl J* 1999 Jan;6(1):1-55. [doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)]
57. Tsetsi E, Rains SA. Smartphone internet access and use: Extending the digital divide and usage gap. *Mobile Media Commun* 2017 Jun 13;5(3):239-255. [doi: [10.1177/2050157917708329](https://doi.org/10.1177/2050157917708329)]
58. Lee YJ, Boden-Albala B, Larson E, Wilcox A, Bakken S. Online health information seeking behaviors of Hispanics in New York City: a community-based cross-sectional study. *J Med Internet Res* 2014 Jul 22;16(7):e176 [FREE Full text] [doi: [10.2196/jmir.3499](https://doi.org/10.2196/jmir.3499)] [Medline: [25092120](https://pubmed.ncbi.nlm.nih.gov/25092120/)]
59. Chang BL, Bakken S, Brown SS, Houston TK, Kreps GL, Kukafka R, et al. Bridging the digital divide: reaching vulnerable populations. *J Am Med Inform Assoc* 2004 Nov;11(6):448-457 [FREE Full text] [doi: [10.1197/jamia.M1535](https://doi.org/10.1197/jamia.M1535)] [Medline: [15299002](https://pubmed.ncbi.nlm.nih.gov/15299002/)]
60. Richardson A, Allen JA, Xiao H, Vallone D. Effects of race/ethnicity and socioeconomic status on health information-seeking, confidence, and trust. *J Health Care Poor Underserved* 2012 Nov;23(4):1477-1493. [doi: [10.1353/hpu.2012.0181](https://doi.org/10.1353/hpu.2012.0181)] [Medline: [23698662](https://pubmed.ncbi.nlm.nih.gov/23698662/)]
61. Ye J, Xu Z, Adesunloye B. Cancer information seeking among black and white adults: results from the 2003 and 2005 health information national trends surveys. *Soc Market Q* 2009 Jul 15;15(1_suppl):16-29. [doi: [10.1080/15245000903038290](https://doi.org/10.1080/15245000903038290)]
62. Perrin A, Duggan M. Americans' internet access: 2000-2015. Pew Research Center. 2015. URL: <https://www.pewresearch.org/internet/2015/06/26/americans-internet-access-2000-2015/#internet-usage-by-gender> [accessed 2021-02-10]
63. Finney RL, Hesse BW, Moser RP, Ortiz MA, Kornfeld J, Vanderpool RC, et al. Socioeconomic and geographic disparities in health information seeking and internet use in Puerto Rico. *J Med Internet Res* 2012 Jul;14(4):e104 [FREE Full text] [doi: [10.2196/jmir.2007](https://doi.org/10.2196/jmir.2007)] [Medline: [22849971](https://pubmed.ncbi.nlm.nih.gov/22849971/)]
64. Bowleg L. The problem with the phrase intersectionality—an important theoretical framework for public health. *Am J Public Health* 2012 Jul;102(7):1267-1273. [doi: [10.2105/ajph.2012.300750](https://doi.org/10.2105/ajph.2012.300750)]
65. Clayman ML, Manganello JA, Viswanath K, Hesse BW, Arora NK. Providing health messages to Hispanics/Latinos: understanding the importance of language, trust in health information sources, and media use. *J Health Commun* 2010 Dec;15 Suppl 3:252-263 [FREE Full text] [doi: [10.1080/10810730.2010.522697](https://doi.org/10.1080/10810730.2010.522697)] [Medline: [21154097](https://pubmed.ncbi.nlm.nih.gov/21154097/)]
66. Mossberger K, Tolbert CJ, Anderson C. The mobile internet and digital citizenship in African-American and Latino communities. *Inform Commun Society* 2016 Oct 17;20(10):1587-1606. [doi: [10.1080/1369118x.2016.1243142](https://doi.org/10.1080/1369118x.2016.1243142)]
67. Heard E, Fitzgerald L, Wigginton B, Mutch A. Applying intersectionality theory in health promotion research and practice. *Health Promot Int* 2020 Aug 01;35(4):866-876. [doi: [10.1093/heapro/daz080](https://doi.org/10.1093/heapro/daz080)] [Medline: [31390472](https://pubmed.ncbi.nlm.nih.gov/31390472/)]
68. Johnson JD, Kreps GL. *Cancer-Related Information Seeking*. New Jersey: Hampton Press; 1997.
69. Johnson J, Meischke H. A comprehensive model of cancer-related information seeking applied to magazines. *Human Comm Res* 1993 Mar;19(3):343-367. [doi: [10.1111/j.1468-2958.1993.tb00305.x](https://doi.org/10.1111/j.1468-2958.1993.tb00305.x)]

Abbreviations

CFI: comparative fit index
HINTS: Health Information National Trends Survey
OHIS: online health information seeking
OR: odds ratio
RMSEA: root mean square error of approximation
RQ: research question
SRMR: standardized root mean square residual

Edited by R Kukafka; submitted 05.08.21; peer-reviewed by M Salimi, JS Tham; comments to author 06.09.21; revised version received 31.10.21; accepted 15.11.21; published 15.03.22.

Please cite as:

Medero K, Merrill Jr K, Ross MQ

Modeling Access Across the Digital Divide for Intersectional Groups Seeking Web-Based Health Information: National Survey

J Med Internet Res 2022;24(3):e32678

URL: <https://www.jmir.org/2022/3/e32678>

doi: [10.2196/32678](https://doi.org/10.2196/32678)

PMID: [35289761](https://pubmed.ncbi.nlm.nih.gov/35289761/)

©Kristina Medero, Kelly Merrill Jr, Morgan Quinn Ross. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 15.03.2022. 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.

Original Paper

Enabling Eating Detection in a Free-living Environment: Integrative Engineering and Machine Learning Study

Bo Zhang¹, DPhil; Kaiwen Deng², MSc; Jie Shen¹, DPhil; Lingrui Cai³, MSc; Bohdana Ratitch¹, DPhil; Haoda Fu¹, DPhil; Yuanfang Guan², DPhil

¹Eli Lilly and Company, Indianapolis, IN, United States

²University of Michigan, Ann Arbor, MI, United States

³Ann Arbor Algorithms, Ann Arbor, MI, United States

Corresponding Author:

Yuanfang Guan, DPhil

University of Michigan

2044D Palmer Commons

Ann Arbor, MI, 48109

United States

Phone: 1 7347440018

Email: gyuanfan@umich.edu

Abstract

Background: Monitoring eating is central to the care of many conditions such as diabetes, eating disorders, heart diseases, and dementia. However, automatic tracking of eating in a free-living environment remains a challenge because of the lack of a mature system and large-scale, reliable training set.

Objective: This study aims to fill in this gap by an integrative engineering and machine learning effort and conducting a large-scale study in terms of monitoring hours on wearable-based eating detection.

Methods: This prospective, longitudinal, passively collected study, covering 3828 hours of records, was made possible by programming a digital system that streams diary, accelerometer, and gyroscope data from Apple Watches to iPhones and then transfers the data to the cloud.

Results: On the basis of this data collection, we developed deep learning models leveraging spatial and time augmentation and inferring eating at an area under the curve (AUC) of 0.825 within 5 minutes in the general population. In addition, the longitudinal follow-up of the study design encouraged us to develop personalized models that detect eating behavior at an AUC of 0.872. When aggregated to individual meals, the AUC is 0.951. We then prospectively collected an independent validation cohort in a different season of the year and validated the robustness of the models (0.941 for meal-level aggregation).

Conclusions: The accuracy of this model and the data streaming platform promises immediate deployment for monitoring eating in applications such as diabetic integrative care.

(*J Med Internet Res* 2022;24(3):e27934) doi:[10.2196/27934](https://doi.org/10.2196/27934)

KEYWORDS

deep learning; eating; digital watch

Introduction

Background

The technological progress of wearable devices, such as smartwatches and wristbands, has made them an integral part of our lives [1]. Wearable devices provide rich, high-frequency, and longitudinal information for symptoms or activities relevant to improving patient diagnosis, care, and treatment. Being able to identify specific relevant activities, such as food intake, in a

way that places a minimal burden on that person has the potential of increasing efficiency of monitoring and patient satisfaction. For example, current diabetes management using basal and bolus insulin regimens requires a high level of patient engagement. One-third of patients with type 1 or type 2 diabetes report insulin omission or nonadherence at least once in the past month, and one of the cited reasons is being too busy [2]. In this scenario, passively collected digital sensor data from consumer wearable devices could be an ideal approach for supplementing the sensor and patient-provided data collected

by specialized connected care diabetes devices. Apart from diabetes, a variety of diseases have been linked to poor eating habits, including heart diseases, obesity, high blood pressure, and other leading causes of death [3,4]. The ability to monitor eating behavior on a continuous basis is central to improving the care and treatment of these conditions.

Related Works

The current literature includes studies of automatic food intake detection using a variety of sensors (Table 1), such as audio, motion, and specialized sensors for chewing and swallowing detection, mounted on different parts of the body such as wrists, head, ears, and neck [5-11]. Although published results are encouraging and indicate the feasibility of automatic food intake detection, advancement in data collection and analytics is still in need. First, most of the existing studies have been conducted in the laboratory [12-15], whereas data on eating in the

free-living environment is more difficult to obtain and infer. Second, if a study is conducted in a free-living condition, it is challenging to obtain accurate ground truth. Typically, such ground truth is obtained through food diaries or questionnaires, and the failure to memorize eating times impedes establishing accurate models [16,17]. Third, because of the cost of wearable watches, participant recruitment, and data extraction, pioneering studies so far are very limited in size, typically covering dozens to hundreds of hours of records in total (Table 1) [18]. For example, Farooq and Sazonov [19] took a total of 23 hours of records >10 individuals in a free-living environment to study the effectiveness of accelerometers in detecting eating. A study that is comparable in size to this one is the Sharma et al [9] study, which contained 1413 hours of records. Finally, this study distinguishes itself from the above studies by its longitudinal follow-up of weeks. This allowed us to update the models for each device user as the data collection proceeded.

Table 1. Representative literature with relatively large size of data on eating detection.

Study	Definitions of eating	Device position	Number of participants	Total hours	F1 score (%)	Weighted accuracy (%)
Dong et al [7]	<ul style="list-style-type: none"> Daily meals and snacks 	Wrist	43	449	N/A ^a	81
Thomaz et al [8]	<ul style="list-style-type: none"> Laboratory: participants were asked to use a fork, knife, hand, and spoon to eat lasagna, popcorn, sandwich, breakfast cereal, rice, and bean Free-living: normal daily meal activities 	Wrist	8	784.25	76.1 and 71.3	N/A
Sharma et al [9]	<ul style="list-style-type: none"> A complete meal or snack 	Wrist	104	1413	N/A	75
Zhang and Amft [20]	<ul style="list-style-type: none"> Participants had no constraints on diet selection and daily activities. They were asked to manually log every eating event in a diet journal of a 1-minute resolution. 	Eyeglasses	10	122.3	95.2 ^b	N/A
Bi et al [11]	<ul style="list-style-type: none"> Laboratory: 6 types of food with 3 crunchy types and 3 soft types Free-living: daily meal activities 	Ear	14	32.2	77.5	92.8
Zhang et al [10]	<ul style="list-style-type: none"> An aggregate of chewing sequences that occur within a short duration of time; these chewing sequences are separated from other chewing sequences by a large time gap 	Neck	20	370.1	81.6	N/A
Farooq et al [19]	N/A	Eyeglasses	10	23	87.9	N/A
This work						
5-minute chunks	N/A	Wrist	34	3828.25	93.8	78
Whole meals in the discovery cohort	N/A	Wrist	34	3828.25	87.7	88
Whole meals in the validation cohort	N/A	Wrist	34	3828.25	87	87

^aN/A: not available.

^bBest.

Objective

Our objective is to develop a prospective, noninterventive, observational study that addresses the above challenges in detecting events of food intake based on passively collected motion sensor data from wearable devices in free-living conditions. We also aim to test the performance of the deep learning algorithms in detecting eating using this data. To this end, we developed a specialized app that allows the recording of eating diaries by simply tapping on the smartwatch and automatic streaming of the accelerometer and gyroscope data into the cloud computing platform. A total record of 3828.25 hours (1658.98 in the discovery cohort and 2169.27 in the validation cohort), encompassing 6 types of eating utensils (forks, knives, spoons, glass, chopsticks, and hands), provided us with deep data for developing models that infer eating behavior in the general population. We develop models that have an area under the curve (AUC) of 0.951 for detecting an entire meal event. We also show the potential to fine-tune more accurate personalized models. A prospective, independent cohort further validated the model. The accuracy of this model supports its immediate readiness to be deployed in clinical trials such as connected diabetes care devices and other therapeutic areas.

Methods

Recruitment and Ethics Approvals

The inclusion criteria of participants in the study were as follows: (1) aged ≥ 18 years; (2) living in the United States; (3) an Eli Lilly employee working in a Lilly office in Indianapolis,

United States; (4) willing to wear an Apple Watch, which is provided for this study and which will be used to collect data from the device motion sensors and logs of events of food consumption; (5) owning a Lilly iPhone and willing to pair it with the Apple Watch provided in this study and to use an app developed for this study to facilitate transfers of motion sensor data; (6) having an internet connection with access to a secure password-protected Wi-Fi at home for the duration of the study; and (7) willing to not use another wrist-worn personal device (eg, Apple Watch) for the duration of this study. The exclusion criteria were as follows: (1) experiencing from hand tremors or involuntary arm movements, (2) currently being a smoker, (3) participation in any other study involving wearable devices that may interfere with the conduct of this study at any point during participation in this study, and (4) being involved in the planning or conduct of this study or being a member of the Machine Learning and Artificial Intelligence team of the Advanced Analytics and Data Sciences group at Eli Lilly. The study has been approved by the Eli Lilly Review Board (study number: 2019-8193) and reviewed by the Western Institutional Review Board (WIRB protocol number 20190878), and all participants have provided written consent to this study. The informed consent form is provided in [Multimedia Appendix 1](#).

Instruments

The purpose of this study is to develop a data streaming system and algorithms that could automatically collect and detect eating events based on passive monitoring of motion sensor data from wearable devices in free-living conditions. The terms used throughout this paper are outlined in [Textbox 1](#).

Textbox 1. Terminology and notations used in the paper.

Terms and explanations	
• Window	• A segment from the data used as the model input (typically 5 minutes in this study)
• Moving step	• The size of the stride between 2 consecutive windows
• Session	• A session is a consecutive recording from the watch; a single day can have multiple sessions.
• Region	• A segment of data within a specific time range
• Aggregation	• The methods that we use to determine the inference of a region based on its related windows
• n	• A cutoff helping to determine if a meal region is inferred correctly or if a region is false positive
• False positive regions	• A region containing at least n false positive data windows
• (Hourly) false positive detection rate	• (Number of the false regions–number of the positive regions)/total sample hours

Our goal is to detect eating activity based on data from motion sensors embedded in wearable devices to minimize the risk of privacy invasion. Eating activity in humans involves potentially distinguishable movements—hand-to-mouth gestures. A total of two motion sensors—accelerometer and gyroscope—were used for the position and orientation sensing in digital watches (Figure 1B). We used the Apple Watch Series 4, which is equipped with both sensors. We programmed the watch to extract sensor data using a standard application programming interface (API), which can be seamlessly paired with an iPhone to facilitate data flow and retrospective labeling. This study has been approved by the Eli Lilly Western Institutional Review Board, and participants' information was deidentified before analytic research. Eating dairy is in general recorded by 2 simple tappings of *begin* and *end* on the Apple Watches. Compared with previous eating diaries, this method facilitates accurate recordings of the eating region.

Participants were asked to log all events of food intake regardless of the type of food or beverage consumed or the manner in which it is consumed (eg, with or without utensils, sitting down, standing, or walking), except while driving. Specifically, they were asked to log each region of food intake if she or he estimated that it would involve >3 bites or sips (movements to bring the food to one's mouth) and would last for >2 minutes. Activities such as taking oral medications, using chewing gum, or taking a few sips of water did not need to be logged. Although many people wear a watch on a nondominant arm, in this study, we asked the participants to wear the study Apple Watch on the arm that they consider dominant for eating purposes. This choice is motivated by the available literature indicating that the food intake detection algorithms using motion sensor data from the dominant arm provide better performance than those using data from the nondominant arm, whereas using sensor data from both arms does not improve performance significantly [5].

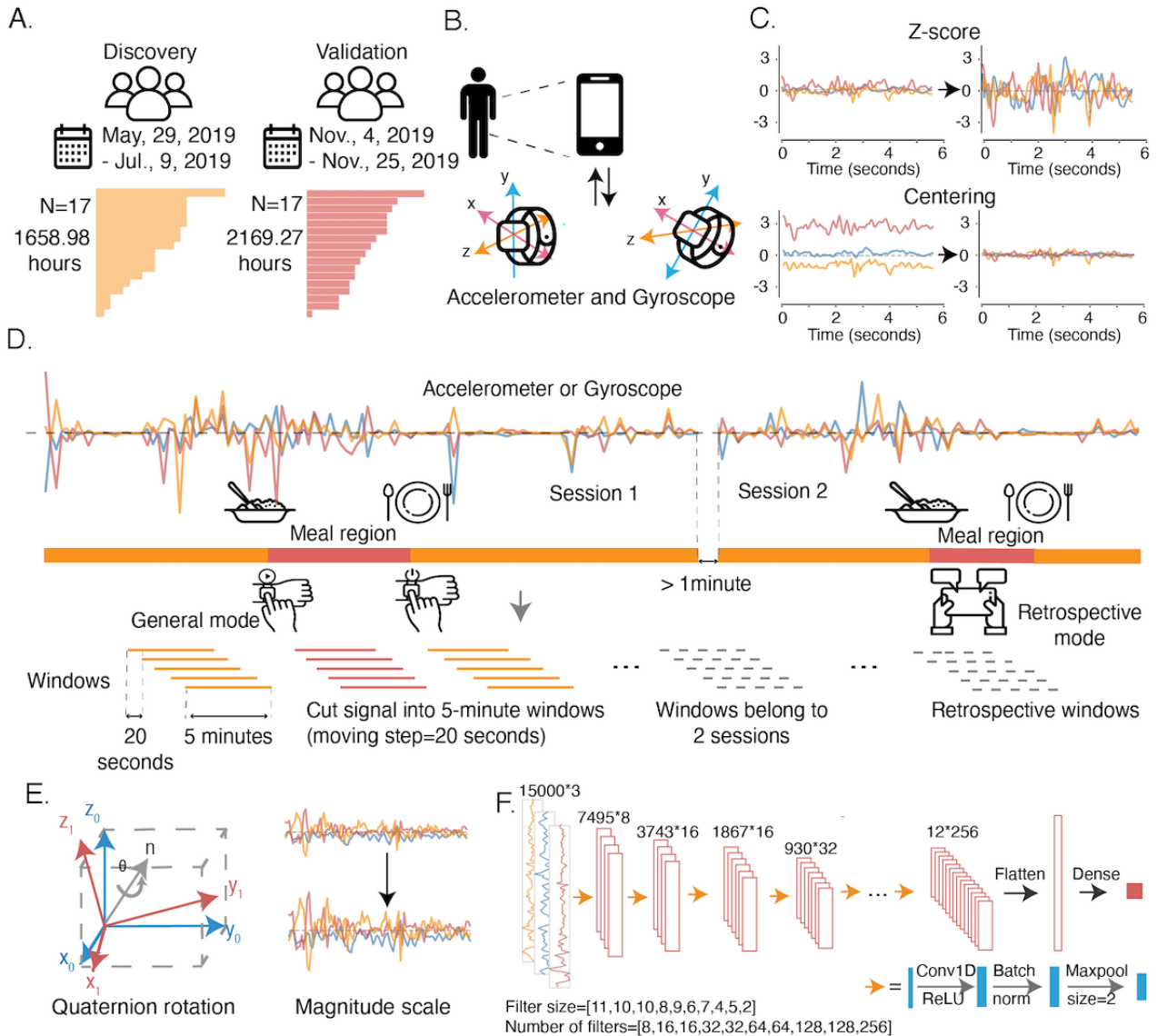
We recruited 2 independent cohorts. The first cohort included 17 individuals, deidentified before data analysis—02CE, 064F, 08A5, 0D51, 0FA7, 11FD, 1453, 16A9, 1B55, 2257, 2BAF, 305B, 32B1, 375 D, 3e5f, 4561, and 47B7—and spanned between May 29, 2019, and July 7, 2019. The second cohort included 17 individuals—766F, 7D71, 7FC7, 8473, 94CD, 9979, A07B, AE7F, B0D5, BED9, C385, CA87, D189, D3DF, DF8D, E1E3, and E68F—and spanned between November 4, 2019,

and November 25, 2019. For each participant, we longitudinally collected a maximum of 20 (discovery cohort) and 22 days (validation cohort) of their daytime activities, with a median of 9 and 11 days, respectively. This provided a total of 1658.98 hours of data in the discovery cohort and 2169.27 hours of data in the validation cohort (Figure 1A). The discovery cohort included 162 days of samples in total, where each individual was allowed to take different numbers of days of experiments varying from 1 (eg, participant 1453) to 20 (eg, participant 2257; Multimedia Appendix 2, Figure S1). The validation cohort included 193 days of samples in total, with the experiment days varying from 1 (eg, participant A07B) to 22 (eg, participant CA87; Multimedia Appendix 2, Figure S1).

Data were recorded at a frequency of 50 Hz and were segmented into individual files by the combination of collection date and participant ID. It is common to see multiple sessions in a single file (Figure 1D and Textbox 1), corresponding to consecutive recording periods in a single day. Approximately 25.3% (41/162) of the samples contain >1 session in the discovery cohort, and approximately 68.9% (133/193) of the samples in the validation cohort have at least two sessions (Multimedia Appendix 2, Figure S1C). The data are presented in a timewise fashion of 20 features, including acceleration and rotation rate ($accl_x$ for acceleration at the x-axis, $accl_y$ for acceleration at the y-axis, $accl_z$ for acceleration at the z-axis, $gyro_x$ for gyroscope at the x-axis, $gyro_y$ for gyroscope at the y-axis, and $gyro_z$ for gyroscope at the z-axis), utensils (binary labels of utensils, fork, knife, spoon, glass, chopstick, and hand), ground truth labels (*tag* for all eating tags, *tagTimely* for eating tags that are done when eating happens, and *tagRetro* for retrospectively recorded taggings), session (*sesid*), timestamp (*ts*), and the local time (*tod*). We determined whether a positive tag should be considered in the training by *tagTimely*, a binary feature indicating whether the tag is labeled during mealtime.

The data collection platform will also enable the participants to retrospectively log approximate times of meals if they forget to log them in a timely manner. The choice of collecting ground truth classification labels through participants' logs is also motivated by the fact that a potential future activity detection system deployed in real life may collect some amount of personalized training data to fine-tune the inference model to individual characteristics.

Figure 1. Overview of data collection and streaming for meal activity analysis. (A) Data comes from two cohorts: 17 participants in the discovery cohort with 1658.98 hours of data and 17 participants in the validation cohorts with 2169.27 hours of data. (B) Signals were collected by 2 sets of sensors, accelerometer and gyroscope, in Apple Watch and paired with iPhone. (C) Z score and centering normalization were conducted for each of the x, y, and z axes for each window. (D) The gyroscope and accelerometer provide continuous signals on the x, y, and z axes over time. There are 2 modes to record meal time. The general model recording starts and stops by tapping the button on the Apple Watch. Retrospective mode allows the participant to type in the rough meal time with the iPhone after having a meal. For each record, we took 5-minute windows with a moving step of 20 seconds. The windows with >2.5 minutes of mealtime will be labeled as eating activity. Windows belonging to 2 sessions are removed. (E) Two augmentation methods, quaternion rotation and scaling the signal magnitude, apply to each data window. (F) Deep learning network structure.



Data Cleaning

A couple of noncompliances appeared to have come from the misunderstandings of the guidance. For example, participant 4561 presented very short sampling regions on June 6, 2019 (Multimedia Appendix 2, Figure S2A). It appears that she/he recorded only the mealtime. Another noncompliance was observed in the close-to-zero signals in accelerometer and gyroscope data for a long region of time; for example, 47B7_2019-06-07 (Multimedia Appendix 2, Figure S2B). This identification number follows the format of participant ID_date. It is likely that participants took off their watch during these time regions. We further removed individuals or days without eating records (eg, 2257_2019-06-27). Communications with participants indicated that they were incorrectly annotated.

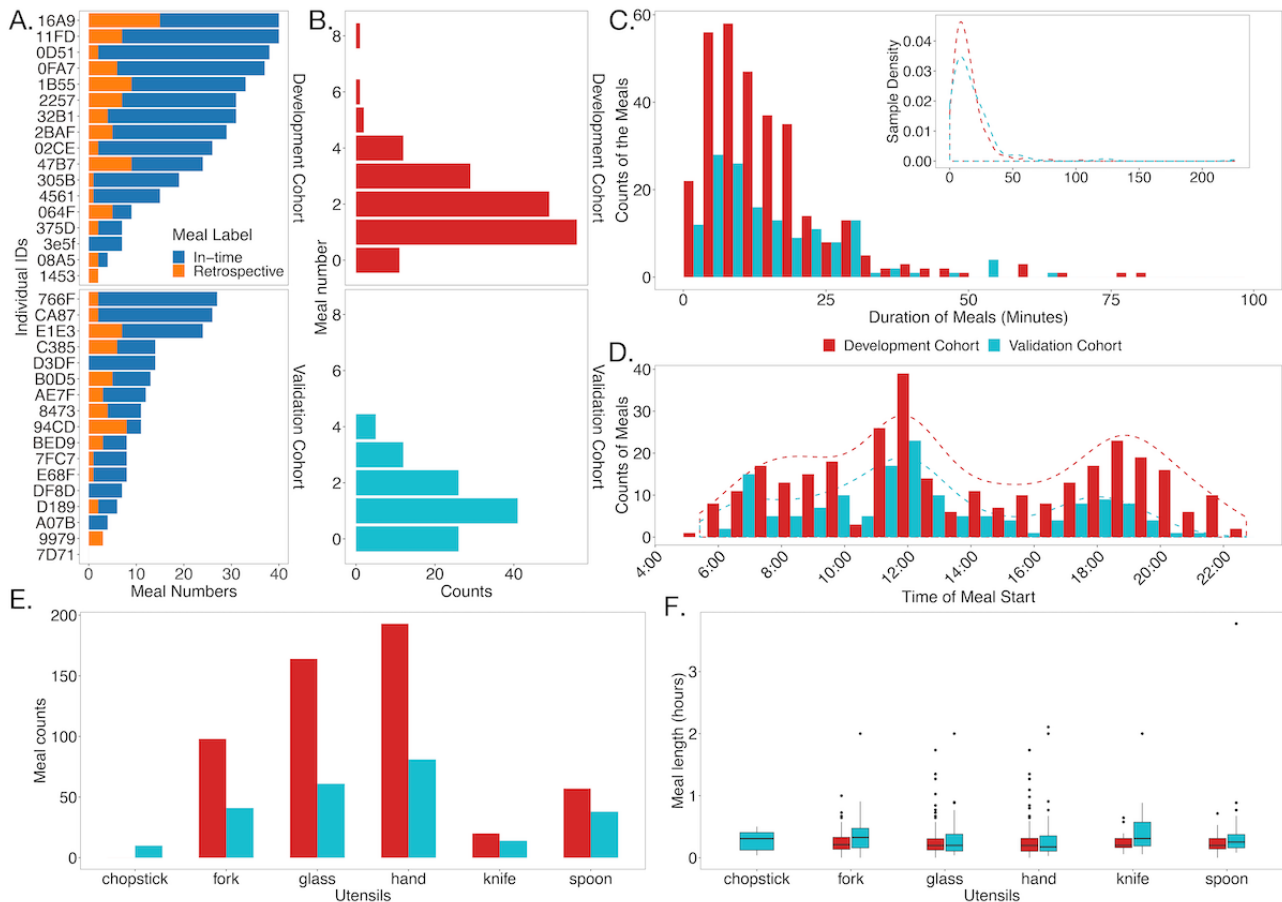
For the discovery cohort, participants 064F, 1453, and 08A5 were excluded from this study because of an overwhelming number of retrospectively annotated meals (>50%; Figure 2A), which indicates potential poor data quality of these days. Records from these participants are also removed as the sampling times are <3 hours in a day, as they are possibly not compliant with the instructions that require wearing the watch during day time: 4561 (May 30, 2019, and June 6, 2019), 305B (June 11, 2019), 47B7 (June 7, 2019, and June 1, 2019), 2257 (June 27, 2019), 375 D (June 7, 2019), 32B1 (June 18, 2019), and 0D51 (June 11, 2019). Of the 17 individuals, 14 (82%) individuals remained to consider in the model development for the global model. The personalized models were fine-tuned and evaluated on the 86% (12/14) of individuals with ≥7 samples:

02CE, 0D51, 0FA7, 11FD, 16A9, 1B55, 2257, 32B1, 2BAF, 305B, 4561, and 47B7.

For the validation cohort, based on the same exclusion rules, we also kept 82% (14/17) of individuals and removed

participants 9979 (3/3); 94CD (8/11), who had an overwhelming number (>50%) of retrospectively recorded meals; and 7D71 (0/0), who failed to label the meal times in all dates. Then, we removed 15 person days with a total duration of <3 hours and another 55 person days, which did not contain any meal times.

Figure 2. Summary of the meal regions. (A) Describes the meal number of each individual in the discovery cohorts and the validation cohorts. The bars are sorted by the total number of the meals and comprise two types of model labeling: label during eating (in-time, the blue bars) and label by retrospect (retrospective, the orange bars). (B) Summarizes the meal numbers per day excluding the retrospectives. (C) Shows the distributions of the duration (in minutes) of the meals for the discovery and the validation cohorts, excluding the retrospective meals. (D) Plots the distributions of the starting times of the meals in these 2 cohorts, and the dashed lines correspond to smoothed curves describing the counts. (E) Distribution of the numbers of the meals with different utensils in the discovery and the validation cohorts. (F) Statistics of the meal lengths (in hours) for different utensils in the discovery and the validation cohorts.



Data Preprocessing

All the data were cut into 5-minute (300 seconds) windows with a moving step of 20 seconds from the start of each date of data based on the record length. The label for each 5-minute data window was determined by the proportion of mealtimes: if the window has >2.5 minutes (150 seconds) labeled as mealtime, then the label of the window is 1 (positive examples); otherwise, it is 0 (negative examples). There were 3 additional conditions that generated -1 labels, which were excluded in both training and evaluation (Figure 1D). The first one is when the current window belonged to 2 different sessions; the second is that the window included the records whose tagRetro (retrospectively recorded eating) was not 0 or missing, which means these tags were recalled by the users after their meals. Third, extremely short periods of eating <3 minutes were excluded as they could have disrupted the fairness of evaluation. A total of 282,942 windows were generated according to this preparation method

from the discovery cohort, and 13,498 of them were positive. As the data were highly unbalanced, we applied an oversampling: we randomly selected N records from the positive examples with replacement, where the N is the number difference between the negative examples and the positive examples.

Model Training and Evaluation

The general training and evaluation strategy was cross-validation, a commonly used scheme that ensures sufficient test examples. In each test, we randomly selected 21% (3/14) of individuals for the test set, 18% (2/11) of individuals for the validation set, and the rest 64% (9/14) of individuals as the training set. Models were trained and tuned on the training and the validation sets, and we evaluated the performances on the test set. We also trained 5 models for each test, which came from 5 random splits on the training set and the validation set while maintaining the 3 final test individuals unchanged. The

final inference scores for the evaluations were the averaged ensemble of the 5 models.

We experimented separately based on accelerometer and gyroscope data, and then we assembled the inference scores using the 2 types of data by taking the average. Experiments are organized in the following order: input data, normalization methods, and augmentations. In each step, we selected the best-performing model for the next experiment. For the fine-tuned personalized model, we first trained a model using all the data in the discovery set, excluding the individual that was the target of fine-tuning. Then, we fine-tuned the model for 2 additional epochs on 60% of the days of the target individual, using another 20% of the days as validation and the last 20% of the days as the test set. Across all experiments, the evaluation was conducted on the testing set with the original class imbalance.

Model Architecture

The backbone of the models is a deep convolutional neural network, comprising 10 building blocks and a fully connected layer for the output listed in [Multimedia Appendix 2](#), Table S13 and [Figure 1F](#). Each block contains a convolutional layer, a batch normalization layer, and a maxpooling layer. The number of filters grows progressively from 8 to 256 (8,16,16,32,32,64,64,128,128, and 256). The sizes of the filters follow (11,10,10,8,9,6,7,4,5, and 2). The network receives both the 3-channel inputs from the accelerometer or the gyroscope and the 6-channel inputs when using them together. The weights of the network are trained by an Adam optimizer [21], the most popular parameter optimizer, with a learning rate of 0.00003 and a binary cross-entropy loss function as the training target is binary. To combat overfitting, we applied a callback function to retrieve the weights from the epoch of the best performance on the validation set. These selected weights were then applied to the test set to evaluate the model performance. We trained a total of 5 epochs. The kernel was initialized with Glorot uniform. The abovementioned parameters were selected empirically and then searched around the empirical values.

Normalization and Data Augmentation

We tested two normalization methods: centering and z score normalization ([Figure 1C](#)). The outputs of centering were the subtraction between the original values and the averages, and the z score normalization required a calculation based on the following formula, where the μ is the average, and the σ is the SD.



To combat overfitting, we applied 2 augmentation methods ([Figure 1E](#)). The first was scaling the signal magnitude by multiplying a randomly selected number from a uniform distribution in (0.8, 1.2). The second was rotating the signals by multiplying a quaternion rotation matrix, which mimics the situations where the same record is taken in different reference frames. First, we randomly generated a set of coordinates (x, y, and z) and defined a reference frame by calculating the basis vectors.



Then, we randomly seeded a rotation angle from $(0, 2\pi)$, and calculated as follows:



The rotation matrix, which multiplies to the original acceleration or gyroscope signals, were then defined as follows:



Evaluation Metrics

The model performances were evaluated on the ensembled inference scores and by a series of metrics, including the area under the receiver operating characteristics curve and the area under the precision–recall curve (AUPRC). Using the information of true positive TP , true negative TN , false positive FP , and false negative FN , we evaluated the weighted F_1 score and the relating precision and recall scores [22], where i is the index of the class, and the w_i =number of class i samples/total sample numbers is the proportion of the class i .



Precisions came from $TP/(TP+FP)$, and recalls came from $TP/(TP+FN)$. We also calculated weighted accuracy following the method in the studies by Dong et al [7] and Sharma et al [9]:



where w is the ratio of the number of negatives over the number of positives.

All the metrics were calculated by the corresponding functions in scikit-learn.

Comparison With DeepConvLSTM

We applied DeepConvLSTM [22] based on the official Pytorch implementation. We used a filter size of 32, and the number of hidden units in the long short-term memory was 64. Details of the structure and the parameters are listed in [Multimedia Appendix 2](#), Table S14. We used the Adam optimizer and binary cross-entropy loss function.

Statistical Significance Analysis

For model comparison, in each test of inference, we first calculated the ratio (denoted as R) of the positive (label=1) to the negative examples (label=0) and then randomly selected 1500 positive 5-minute windows and $1500R$ negative windows. We repeated 100 times to estimate the P values.

Code Availability

The code is attached with the submission ([Multimedia Appendix 3](#)) and can be runnable with Python 3.6.12 and Keras 2.2.4.

Data Availability

On the basis of consent forms, Eli Lilly can share data with regulatory authorities (Food and Drug Administration) in the

United States, the ethical review board overseeing this study, and the researchers at other institutions who wish to analyze the data in this study.

Results

Deep Learning Accurately Classifies Eating Activity in 5-Minute Windows on Previously Unseen Individuals

We first analyzed the data from the discovery cohort. Raw data were collected from the accelerometer and gyroscope from the Apple Watch at a frequency of 50 Hz and streamed to Amazon Web Service [23]. Each time point was labeled with a meal tag (1 denoted the meal region and 0 denoted non-meal time). Participants were asked to specify the start and the end of the mealtime and whether this region was recorded at the time of the meal or retrospectively. Most (126/162, 77.8%) of the daily records lasted approximately 8 to 15 hours, representing the daily activity time when the participants wore the watches (Multimedia Appendix 2, Figure S1B). The participants are likely to begin their records from 7 AM to 9 AM (Multimedia Appendix 2, Figure S1D) and end at 7 PM to 9 PM (Multimedia Appendix 2, Figure S1E), which is consistent with the expected daily activity time. A participant could have 1 to 7 eating events within a day, with the vast majority having between 1 and 4 eating events per day. Approximately 75% of the meals would last for <20 minutes. The start and peak times of the meal events were shown at the expected breakfast, lunch, and dinner times (Figure 2).

With the generated 5-minute windows (Figure 1D), we constructed a 1D (along the time axis for both input and output) deep learning model (Figure 1F) with 3 channels as input (x, y, and z axes of accelerometer or gyroscope; 6 channels when giving both accelerometer and gyroscope information). On the basis of the cross-validation described in the *Methods* section, our work showed an average AUC of 0.825 (SD 0.073; Figure 3D and Figure 4C) and an average AUPRC of 0.437 (SD 0.096), with the baseline (same value predictions for all data points) of 0.053 (Multimedia Appendix 2, Figure S3C and Figure 4D). When including the retrospective meals in prediction, our model showed stable performances with an average AUC of 0.813 (SD 0.067) and AUPRC of 0.440 (SD 0.077, baseline 0.065). In comparison, we adapted DeepConvLSTM on this data set [22], which achieved an average AUC of 0.797 (SD 0.065) and an average AUPRC of 0.294 (SD 0.072; Figure 3E) on the nonretrospective meals. This demonstrated that the techniques integrated into this approach could substantially improve over a state-of-the-field method.

We identified the factors that affect performance. First, based on the 5 models trained on the random splits of the training set, assembling the inference values from the output of the last fully connected layer, by taking the averages in each test, can significantly improve the performances in all the experiments ($P < .001$; Figure 3B-Figure 3D; Multimedia Appendix 2, Figure S3A-3C). Second, building the model on gyroscope data can achieve better performances than using accelerometer data or both. The average AUC and AUPRC of the gyroscope model are 0.02 to 0.05 higher than the other alternatives (P values for AUCs $< .001$; P values for AUPRCs $< .001$; Figure 3B and Multimedia Appendix 2, Figure S3A; Multimedia Appendix 2, Tables S1 and S2). Third, choosing correct input data normalization methods may be helpful. Centering normalization improved the model performance by 0.002 on the AUC and 0.01 on the AUPRC (P values for AUCs=.29; P values for AUPRCs=.10), whereas, with the z score normalization, which may compress the original ranges of the signals, the performances will drop by 0.01 and 0.04 on the AUC and AUPRC (Figure 3C and Multimedia Appendix 2, S3B; Multimedia Appendix 2, Tables S3 and S4). This is likely to reflect the fact that the magnitude of the signal is critical to the model, whereas the directions of the watch (ie, reflected as the overall shift of an axis) are not relevant. Fourth, data augmentation, including the quaternion rotation of the signals in space and scaling the signal magnitudes, might improve the model performance. Rotation and scaling can provide >0.01 improvement on both AUCs and AUPRCs for a single model, although not statistically significant (P values for AUCs=.33; P values for AUPRCs=.17). When considering the ensemble model that aggregates 5 models generated using different random seeds, the magnitude scaling gives better but not significantly better performance both on AUC and AUPRC (P values for AUCs=.26; P values for AUPRCs=.88; Figure 3D and Multimedia Appendix 2, Figure S3C; Multimedia Appendix 2, Tables S3 and S4). Adding in local time did not improve the performance (Tables S1 and S2).

To retrieve the performances of each individual, including the previously excluded ones, and generate the baselines for evaluating the improvements of our following fine-tunings on the personalized models, we also used the leave-one-subject-out approach to calculate the AUCs and AUPRCs. For each individual, the model was trained on all the other data except the one left out. The average AUC for the ensemble model was 0.818 (SD 0.104), and the average AUPRC was 0.419 (SD 0.162, Figure 4A and Figure 4B; Multimedia Appendix 2, Table S7). Visualization of the inferences of two dates of records: data on June 24, 2019, from 0FA7 and data on June 10, 2019, from 3E5F show consistency with the eating and noneating behaviors (Figure 4E and Figure 4F).

Figure 3. Evaluation of model performance on 5-minute windows of the discovery cohort. (A) Models were built by gyroscope data only, accelerometer data only, and gyroscope+accelerometer data. Next, we tested the centering and normalization of each axis of the data. Intensive data augmentation was applied to the data on the fly. For each method, 5 models were trained by resampling the training and validation data and they were assembled for evaluation. (B) Presents the performance comparisons of different data selections. (C) Presents the performance comparison of different normalization methods applied on the gyroscope model. (D) Presents the performance comparisons of the augmentation methods based on the centering model, where Quart refers to the quaternion rotation augmentation, and Scale refers to scaling the magnitude. (E) Comparison of the performances between DeepConvLSTM and the method presented in this paper. AUROC: area under the ROC curve; AUPRC: area under the precision–recall curve; CNN: convolutional neural network; ROC: receiver operator characteristic.

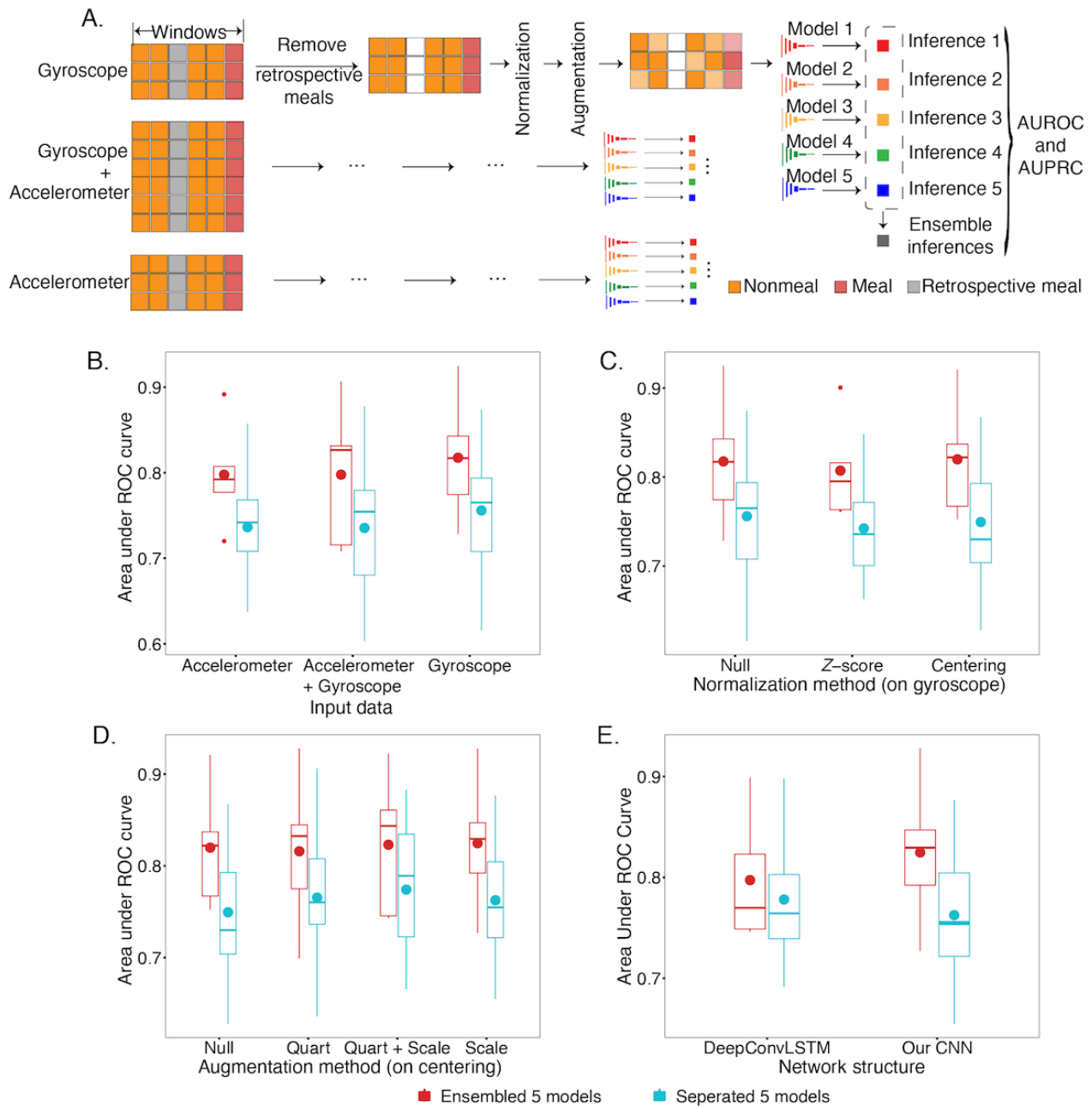
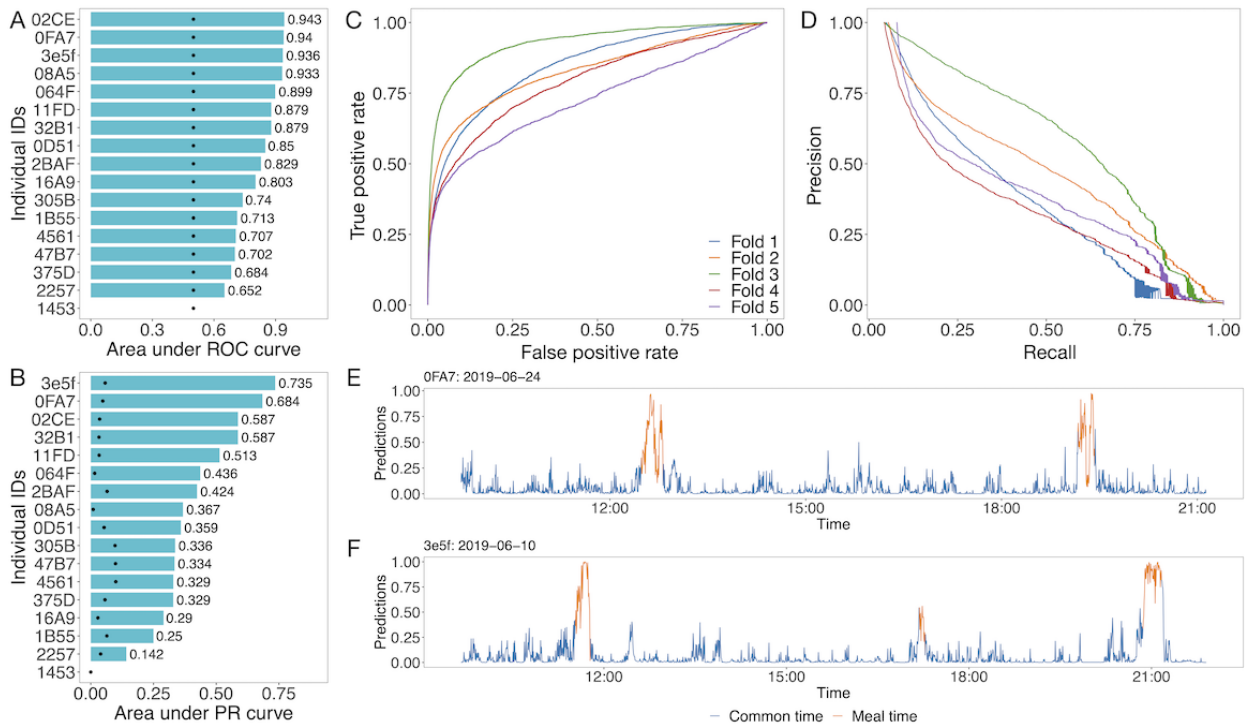


Figure 4. The evaluations of the selected best model: convolutional neural network backbone trained with the centering normalized gyroscope data using magnitude scaling. (A and B) show the leave-one-subject-out results of the model on the discovery cohort, evaluated both in the area under the curves and area under the precision–recall curves. The black points in (B) are the baselines for the individuals. As 1453 does not have any positive samples after excluding the retrospectives, its value will be empty. (C) is an area under the ROC curve, and (D) is a precision–recall curve for the cross-validation from the ensemble model, respectively. (E and F) give the inferences of the 2 dates of records, where the blue segments denote the signals of the non–meal time, and the orange segments are the signals of the mealtime. PR: precision recall; ROC: receiver operator characteristic.

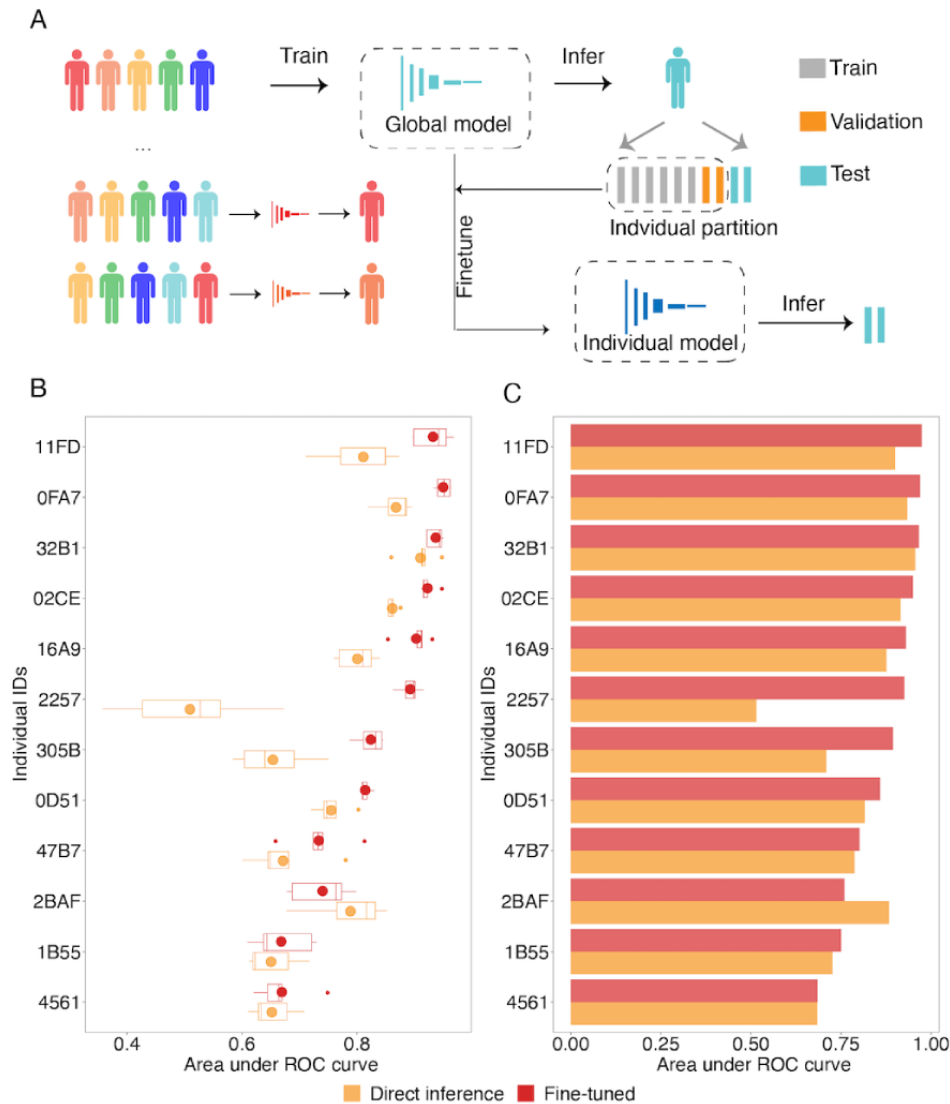


Fine-tuning of the Personalized Model Improves Performance

This longitudinal data allowed us to explore whether it is possible to construct personalized models for eating and further improve model performance (Figure 5A). The global models served as the fine-tuning starting points for the individuals of interest (Figure 4A and Figure 4B). This study design mimics an important utility of the models in real life, where we adopted an existing model to a previously unseen person and asked whether we could improve the inference on this individual by observing some data for this individual.

Comparing the performance of the global model on this individual versus the fine-tuned model, we found that other than 1 individual (2BAF), the fine-tuned personalized models showed better performance than directly applying the population models on the specific individuals. The AUC on average improved for the fine-tuning model to 0.872 (SD 0.099), with an average weighted F1 (the average weights were 0.059 and 0.941 for positives and negatives, respectively) score of 0.938 (SD 0.048), an average precision of 0.945 (SD 0.045), and an average recall of 0.934 (SD 0.049; Figure 5B and Figure 5C; Multimedia Appendix 2, Tables S8 and S9).

Figure 5. Evaluating fine-tuning to generate individual models on the discovery cohort. (A) For a specific individual under investigation, we first trained 5 global models using all other individuals by resampling the training and validation set for the deep learning training process. Next, we split records of the individual of interest by days into training, validation, and test sets and fine-tuned the global model using the training and validation set. We evaluated the performance by the area under the ROC curve for both the global and the individual fine-tuned models for (B) 5 separated models and (C) ensemble models. ROC: receiver operator characteristic.



Aggregation of Multiple Windows Reaches Near-Perfect Detection of Meal Events

We then evaluated the model performance on the original mealtimes. We conducted three experiments on cross-validation of the discovery cohort: (1) the prediction for whole meals (Figure 6A), (2) the prediction within 5 minutes or 10 minutes after the meal starts (Figure 6B), and (3) the false calls within an hour (Figure 6C). For all nonretrospectively recorded meal events, we calculated the average score during each meal event. For calculating scores for the negative regions, we randomly selected a series of negative regions whose lengths and numbers were matched to the meal events. The scores for the negatives were generated by taking the averages of the windows within the selected regions. The models achieved an AUC of 0.951 (SD 0.018) by this aggregation, and the corresponding weighted F1 score (weights were 0.464 and 0.536 for positives and negatives, respectively), precision, and recall were 0.877 (SD 0.037), 0.8890 (SD 0.027), and 0.879 (SD 0.035), respectively

(Figure 6D; Multimedia Appendix 2, Table S10). Including the retrospective, meals would result in a similar AUC of 0.951 (SD 0.017), with the corresponding weighted F1 score of 0.858 (SD 0.040, weights for positives and negatives were 0.5).

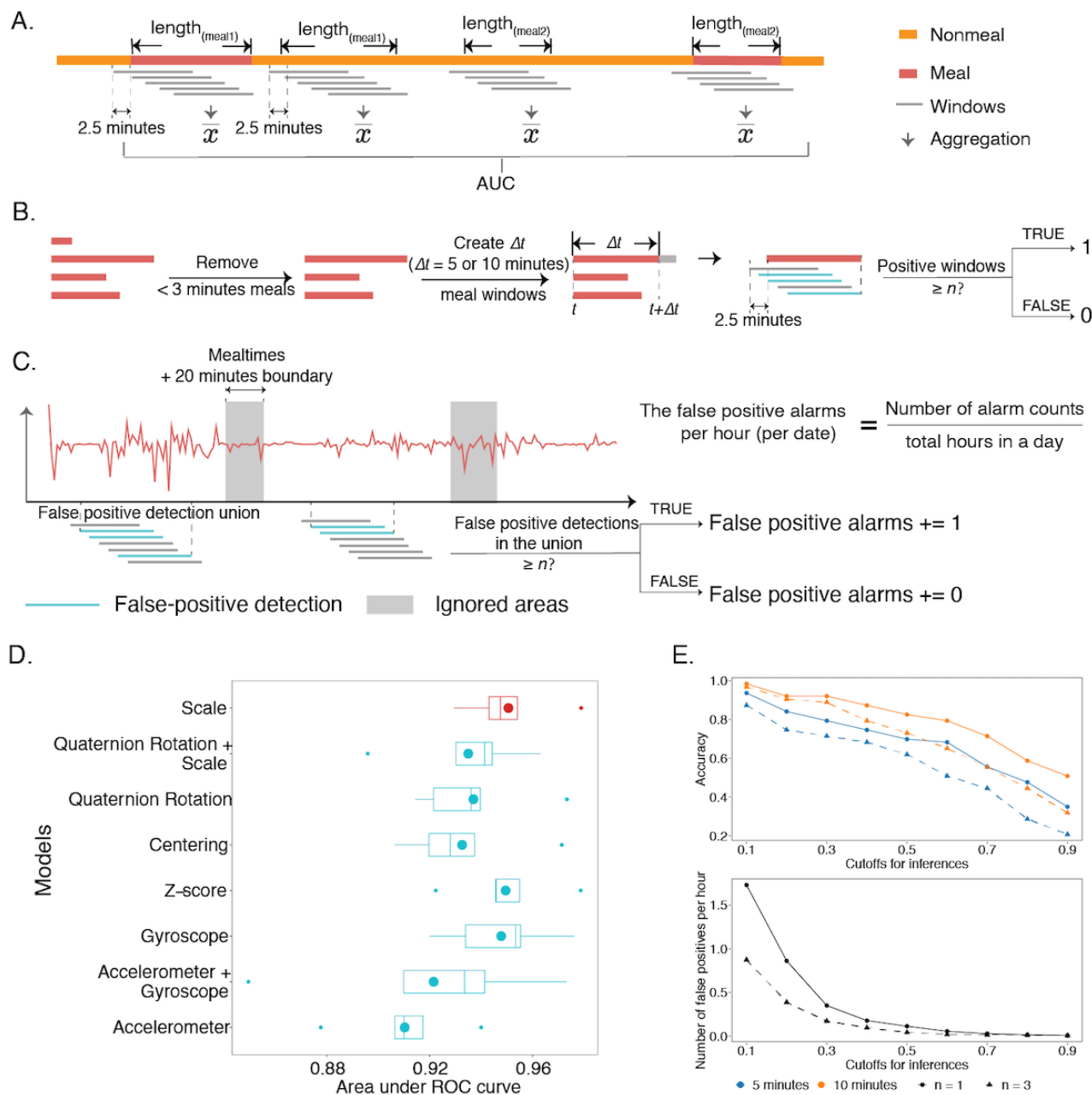
For prediction on the 10-minute or 5-minute after meal start, we used the accuracy (ie, how many mealtimes were correctly inferred) for evaluation (Figure 6E; Multimedia Appendix 2, Tables S11 and S12). In this case, we were interested in how we could choose a criterion so that most meals could be detected within 5/10 minutes. We used a moving step of 10 seconds and defined a window to be positive if the prediction score was >0.3. If >3 windows were predicted to be positive in 5/10 minutes, we alarmed a call. Using these criteria, we reached a recall rate of 0.889 for 10 minutes, and the result remained robust in the 5-minute test, cutoffs of prediction score between 0.4 and 0.6, and the number of windows we used to alarm the call.

Next, we calculated the number of false positive predictions per hour of negative regions. For each hour, we had a total of 360

chunks (with a moving step of 10 seconds). The corresponding false positive prediction was 0.172 per hour; that is, 1 to 2 times of false positives in a whole day of activity. Again, this result

was robust against the cutoff defining a positive window and the number of windows needed to alarm a call.

Figure 6. The results of the aggregations on the entire meals and on the specific time regions after meal start. (A) Aggregation to produce meal-level detection performance. (B) A preprocessing step removed meals that were <3 minutes and created 5- or 10-minute windows for positive examples to evaluate recall rate. (C) Nonmeal regions were used to calculate false positive alarms using the 5-minute windows and the same criteria as what was used to define positive inferences in calculating the recall rate. The gray areas denote the mealtimes with the 10-minute boundaries at the start and the end, where the windows are out of consideration. (S) Shows the evaluations of the aggregations on the entire meals. The boxplots comprise the AUCs from the average inferred scores of the ensemble models in the cross-validation, and the experiments (models) are the same as those in Figure 3A. The point in each box denotes the corresponding average AUC. (E) indicates the results of the aggregation on the 5- and 10-minute meals after starting and on the entire negative signals. The lines show how the detection accuracy and the hourly false positive numbers (the black lines) change along with the cutoffs. The orange lines show the results for the 10-minute meals, and the blue lines are the 5-minute meals. The shape of the points represents the choice of N, where the circles/solid lines are N=1, and the triangles/dashed lines are N=3. AUC: area under the curve; ROC: receiver operator characteristic.



Generalizability to an Independent Validation Cohort Collected in a Different Season

Although the first batch of the data was collected in the summer, we proceeded to collect a second validation cohort in winter, 6

months later, by recruiting 17 new individuals. By splitting the discovery cohort data into 5 sets of training and validation data, we first finalized 5 models for the first cohort; then, we directly applied these models to the validation cohort for inferences. Next, we applied the scheme of the whole meal predictions to

the validation cohort, both on the data with and without the retrospective meals. Without any further tuning, the model achieved a meal-level AUC of 0.941 on the validation cohort for the nonretrospective meals, with a 0.870 weighted F1 score (the weights were 0.445 and 0.555 for positives and negatives,

respectively), a 0.878 precision, and a 0.871 recall. With the retrospective meals, the meal-level AUC and the weighted F1 score were 0.920 and 0.846, respectively (the weights for positives and negatives were 0.5). The performances of our work in this study are listed in [Table 2](#).

Table 2. List of the performances in this study.

Experiments	Area under the curve	Area under the precision–recall curve	Weighted F1 score
Cross-validation of our model on 5-minute windows	0.825	0.437	N/A ^a
Cross-validation of our model on 5-minute windows, including the predictions on retrospective meals	0.813	0.440	N/A
Cross-validation of DeepConvLSTM [22] on 5-minute windows	0.797	0.294	N/A
Leave-one-subject-out approach of our best model on 5-minute windows	0.818	0.419	N/A
Fine-tuning the personalized model	0.872	N/A	0.938
Cross-validation of our model on the original mealtimes (discovery cohort)	0.951	N/A	0.877
Cross-validation of our model on the original mealtimes, including the predictions of retrospective meals (discovery cohort)	0.951	N/A	0.858
Predictions of our model on the original mealtimes (validation cohort)	0.941	N/A	0.870
Predictions of our model on the original mealtimes, including the predictions of retrospective meals (validation cohort)	0.920	N/A	0.846
Accuracy of detecting the eating in 10 minutes	0.889	N/A	N/A
False positive detections per hour	0.172	N/A	N/A

^aN/A: not applicable.

Discussion

Principal Findings

In this study, we presented a large, in-the-field, digital eating detection study of eating activity. Deep learning algorithms experimented with a diverse array of augmentation, preprocessing, and architectures allowed us to narrow down the algorithm into one with a performance of AUC of 0.825 to infer previously unseen individuals for a single 5-minute window. When evaluated on the entire meal regions, this AUC was 0.951. We further validated the algorithm in an independently time-lapsed cohort collected in a different season (6 months later, winter) and achieved a meal-level performance of 0.941 AUC without further tuning. This design can potentially result in models that are more or at least similarly generalizable than data collected consecutively in the same season. This represents the first study that harbors a validation cohort in this field.

Compared with other studies that focus on population-wise models [9,13,24-26], the longitudinal weeks of follow-ups of the data set presented in this study allowed us to further explore the possibility of personalized models for detecting eating activity. It is widely recognized that eating motions differ substantially in a population by gender, culture, and certainly individual habits. This fine-tuning scheme produced an average AUC at 0.872, corresponding to a 0.89 success rate in calling back an eating event within 10 minutes. This substantial improvement in performance points to the direction toward personalized eating monitoring in the dietary research field.

Records of the local time, as well as the utensils used for each meal, also allow us to glean insight into their influences on our model ([Multimedia Appendix 2](#), Figure S4). We found that food taken with hands had relatively poor performance (AUC=0.812; [Multimedia Appendix 2](#), Figure S4B). In addition, we found that false positive rates are relatively high between 6 AM to 7 AM and 9 PM, indicating potential morning and evening activities mimicking eating movement ([Multimedia Appendix 2](#), Figure S4C). Future studies incorporating different characteristics of utensils as well as whole daily activity logs might have the potential to further improve the performance.

Limitations

We acknowledge several potential limitations of this study. First, we excluded smoking individuals, for whom the inference task could become more complicated as the motion of smoking shares a certain similarity with the motion of eating. Second, we only included healthy individuals, which may not be representative of the population of movement disorders such as ataxia and Parkinson disease. In addition, we did not collect the data for the nondominant hand. The weaker and noisy signals may significantly affect our model built on dominant hand data. Potentially, combined with additional devices such as ear- and chest-anchored devices and video ([Multimedia Appendix 2](#), Table S15 [27-31]), in future works, we will be able to combat such limitations. We used a total of 34 individuals in the study. Although we observed strong predictions across individuals, larger collections focusing on more individuals but less longitudinal follow-up might further complement the information provided in this study. Furthermore, we used 50

Hz data in this study for optimizing battery performance in collecting data. It is yet to be evaluated how higher Hz data contribute to performance with the development of the devices.

Future Works and Conclusions

This study and the API developed here open several future directions that are worth exploration. For example, how do digital indicators differ for populations coming from different cultural backgrounds? Does handedness affect model construction and performance? And how much will the model

be affected if one wears the device on his or her nondominant hand? Answering these questions will need large-scale studies with a large number of participants, and the API and data streaming platform developed in this study will become a convenient tool for this purpose. The accuracy of the models developed in this study satisfies immediate deployment needs in clinical settings to monitor eating behavior and give guidance to treatment regimen adjustment accordingly. We envision the digital streaming platform will be widely integrated into a variety of clinical trials in the near future.

Authors' Contributions

BZ, JS, BR, and HF contributed to the study design, device design, and data streaming. KD, JS, YG, and BR contributed to data analytics. YG drafted the manuscript. KD and LC were responsible for the figures. BR contributed to conducting and analysis planning.

Conflicts of Interest

BZ and HF are current Eli Lilly and Company employees. JS and BR were Eli Lilly and Company employees when the work was conducted. KD and LC were Ann Arbor Algorithms, Inc employees when this study was conducted. YG serves as the scientific adviser for Eli Lilly and Company on this study.

Multimedia Appendix 1

The study consent form.

[[DOC File, 83 KB - jmir_v24i3e27934_app1.doc](#)]

Multimedia Appendix 2

The supplementary tables and figures.

[[PDF File \(Adobe PDF File\), 1156 KB - jmir_v24i3e27934_app2.pdf](#)]

Multimedia Appendix 3

The codes for our models and experiments in this study.

[[ZIP File \(Zip Archive\), 363 KB - jmir_v24i3e27934_app3.zip](#)]

References

1. Metcalf D, Milliard ST, Gomez M, Schwartz M. Wearables and the internet of things for health: wearable, interconnected devices promise more efficient and comprehensive health care. *IEEE Pulse* 2016 Sep;7(5):35-39. [doi: [10.1109/mpul.2016.2592260](#)]
2. Peyrot M, Barnett AH, Meneghini LF, Schumm-Draeger P. Insulin adherence behaviours and barriers in the multinational Global Attitudes of Patients and Physicians in Insulin Therapy study. *Diabet Med* 2012 May;29(5):682-689 [[FREE Full text](#)] [doi: [10.1111/j.1464-5491.2012.03605.x](#)] [Medline: [22313123](#)]
3. Pot GK, Hardy R, Stephen AM. Irregularity of energy intake at meals: prospective associations with the metabolic syndrome in adults of the 1946 British birth cohort. *Br J Nutr* 2015 Nov 09;115(2):315-323. [doi: [10.1017/s0007114515004407](#)]
4. Almoosawi S, Vingeliene S, Karagounis LG, Pot GK. Chrono-nutrition: a review of current evidence from observational studies on global trends in time-of-day of energy intake and its association with obesity. *Proc Nutr Soc* 2016 Nov;75(4):487-500. [doi: [10.1017/S0029665116000306](#)] [Medline: [27327252](#)]
5. Mirtchouk M, Lustig D, Smith A, Ching I, Zheng M, Kleinberg S. Recognizing eating from body-worn sensors. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2017 Sep 11;1(3):1-20 [[FREE Full text](#)] [doi: [10.1145/3131894](#)]
6. Fullerton E, Heller B, Munoz-Organero M. Recognizing human activity in free-living using multiple body-worn accelerometers. *IEEE Sensors J* 2017 Aug 15;17(16):5290-5297 [[FREE Full text](#)] [doi: [10.1109/jsen.2017.2722105](#)]
7. Dong Y, Scisco J, Wilson M, Muth E, Hoover A. Detecting periods of eating during free-living by tracking wrist motion. *IEEE J Biomed Health Inform* 2014 Jul;18(4):1253-1260. [doi: [10.1109/jbhi.2013.2282471](#)]
8. Thomaz E, Essa I, Abowd G. A practical approach for recognizing eating moments with wrist-mounted inertial sensing. In: *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, 2015 Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 7 - 11, 2015; Osaka Japan.* [doi: [10.1145/2750858.2807545](#)]

9. Sharma S, Jasper P, Muth E, Hoover A. Automatic detection of periods of eating using wrist motion tracking. In: Proceedings of the IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). 2016 Presented at: IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); Jun 27-29, 2016; Washington, DC, USA. [doi: [10.1109/chase.2016.15](https://doi.org/10.1109/chase.2016.15)]
10. Zhang S, Zhao Y, Nguyen DT, Xu R, Sen S, Hester J, et al. NeckSense: a multi-sensor necklace for detecting eating activities in free-living conditions. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2020 Jun;4(2):72 [FREE Full text] [doi: [10.1145/3397313](https://doi.org/10.1145/3397313)] [Medline: [34222759](https://pubmed.ncbi.nlm.nih.gov/34222759/)]
11. Bi S, Wang T, Tobias N, Nordrum J, Wang S, Halvorsen G, et al. Auracle: detecting eating episodes with an ear-mounted sensor. *Proc ACM Interact Mob Wearable Ubiquitous Technol* 2018 Sep 18;2(3):1-27. [doi: [10.1145/3264902](https://doi.org/10.1145/3264902)]
12. Kyritsis K, Diou C, Delopoulos A. End-to-end learning for measuring in-meal eating behavior from a smartwatch. In: Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2018 Presented at: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 18-21, 2018; Honolulu, HI, USA. [doi: [10.1109/embc.2018.8513627](https://doi.org/10.1109/embc.2018.8513627)]
13. Thomaz E, Bedri A, Prioleau T, Essa I, Abowd G. Exploring symmetric and asymmetric bimanual eating detection with inertial sensors on the wrist. *DigitalBiomarkers* 17 (2017) 2017 Jul;2017:21-26 [FREE Full text] [doi: [10.1145/3089341.3089345](https://doi.org/10.1145/3089341.3089345)] [Medline: [29505038](https://pubmed.ncbi.nlm.nih.gov/29505038/)]
14. Papadopoulos A, Kyritsis K, Sarafis I, Delopoulos A. Personalised meal eating behaviour analysis via semi-supervised learning. In: Proceedings of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2018 Presented at: 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Jul 18-21, 2018; Honolulu, HI, USA. [doi: [10.1109/embc.2018.8513174](https://doi.org/10.1109/embc.2018.8513174)]
15. Cho J, Choi A. Asian-style food intake pattern estimation based on convolutional neural network. In: Proceedings of the IEEE International Conference on Consumer Electronics (ICCE). 2018 Presented at: IEEE International Conference on Consumer Electronics (ICCE); Jan 12-14, 2018; Las Vegas, NV, USA. [doi: [10.1109/icce.2018.8326311](https://doi.org/10.1109/icce.2018.8326311)]
16. Livingstone MB, Robson PJ, Wallace JM. Issues in dietary intake assessment of children and adolescents. *Br J Nutr* 2004 Oct;92 Suppl 2:S213-S222. [doi: [10.1079/bjn20041169](https://doi.org/10.1079/bjn20041169)] [Medline: [15522159](https://pubmed.ncbi.nlm.nih.gov/15522159/)]
17. Westertep KR, Goris AH. Validity of the assessment of dietary intake: problems of misreporting. *Curr Opin Clin Nutr Metab Care* 2002 Sep;5(5):489-493. [doi: [10.1097/00075197-200209000-00006](https://doi.org/10.1097/00075197-200209000-00006)] [Medline: [12172471](https://pubmed.ncbi.nlm.nih.gov/12172471/)]
18. Bell BM, Alam R, Alshurafa N, Thomaz E, Mondol AS, de la Haye K, et al. Automatic, wearable-based, in-field eating detection approaches for public health research: a scoping review. *NPJ Digit Med* 2020 Mar 13;3(1):38 [FREE Full text] [doi: [10.1038/s41746-020-0246-2](https://doi.org/10.1038/s41746-020-0246-2)] [Medline: [32195373](https://pubmed.ncbi.nlm.nih.gov/32195373/)]
19. Farooq M, Sazonov E. Accelerometer-based detection of food intake in free-living individuals. *IEEE Sensors J* 2018 May 1;18(9):3752-3758. [doi: [10.1109/jsen.2018.2813996](https://doi.org/10.1109/jsen.2018.2813996)]
20. Zhang R, Amft O. Free-living eating event spotting using EMG-monitoring eyeglasses. In: Proceedings of the IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). 2018 Presented at: IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); Mar 4-7, 2018; Las Vegas, NV, USA. [doi: [10.1109/bhi.2018.8333386](https://doi.org/10.1109/bhi.2018.8333386)]
21. Zhang Z. Improved Adam optimizer for deep neural networks. In: Proceedings of the IEEE/ACM 26th International Symposium on Quality of Service (IWQoS). 2018 Presented at: IEEE/ACM 26th International Symposium on Quality of Service (IWQoS); Jun 4-6, 2018; Banff, AB, Canada. [doi: [10.1109/iwqos.2018.8624183](https://doi.org/10.1109/iwqos.2018.8624183)]
22. Ordóñez FJ, Roggen D. Deep convolutional and LSTM recurrent neural networks for multimodal wearable activity recognition. *Sensors (Basel)* 2016 Jan 18;16(1):115 [FREE Full text] [doi: [10.3390/s16010115](https://doi.org/10.3390/s16010115)] [Medline: [26797612](https://pubmed.ncbi.nlm.nih.gov/26797612/)]
23. Varia J, Mathew S. Overview of amazon web services. *Amazon Web Services*. 2014. URL: http://cabibbo.dia.uniroma3.it/asw-2014-2015/altrui/AWS_Overview.pdf [accessed 2022-01-31]
24. Farooq M, Sazonov E. Detection of chewing from piezoelectric film sensor signals using ensemble classifiers. In: Proceedings of the 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 16-20, 2016; Orlando, FL, USA. [doi: [10.1109/embc.2016.7591833](https://doi.org/10.1109/embc.2016.7591833)]
25. Kyritsis K, Diou C, Delopoulos A. Food intake detection from inertial sensors using LSTM networks. In: *New Trends in Image Analysis and Processing*. Switzerland: Springer; 2017.
26. Farooq M, Fontana J, Boateng A, Mccrory M, Sazonov E. A comparative study of food intake detection using artificial neural network and support vector machine. In: Proceedings of the 12th International Conference on Machine Learning and Applications. 2013 Presented at: 12th International Conference on Machine Learning and Applications; Dec 4-7, 2013; Miami, FL, USA. [doi: [10.1109/icmla.2013.33](https://doi.org/10.1109/icmla.2013.33)]
27. Nishimura J, Kuroda T. Eating habits monitoring using wireless wearable in-ear microphone. In: Proceedings of the 3rd International Symposium on Wireless Pervasive Computing. 2008 Presented at: 3rd International Symposium on Wireless Pervasive Computing; May 7-9, 2008; Santorini, Greece. [doi: [10.1109/iswpc.2008.4556181](https://doi.org/10.1109/iswpc.2008.4556181)]
28. Ravi D, Lo B, Yang GZ. Real-time food intake classification and energy expenditure estimation on a mobile device. In: Proceedings of the IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN). 2015

- Presented at: IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN); Jun 9-12, 2015; Cambridge, MA, USA. [doi: [10.1109/bsn.2015.7299410](https://doi.org/10.1109/bsn.2015.7299410)]
29. Liu J, Johns E, Atallah L, Pettitt C, Lo B, Frost G, et al. An intelligent food-intake monitoring system using wearable sensors. In: Proceedings of the Ninth International Conference on Wearable and Implantable Body Sensor Networks. 2012 Presented at: Ninth International Conference on Wearable and Implantable Body Sensor Networks; May 9-12, 2012; London, UK. [doi: [10.1109/bsn.2012.11](https://doi.org/10.1109/bsn.2012.11)]
30. Gao Y, Zhang N, Wang H, Ding X, Ye X, Chen G, et al. iHear food: eating detection using commodity bluetooth headsets. In: Proceedings of the IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). 2016 Presented at: IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); Jun 27-29, 2016; Washington, DC, USA. [doi: [10.1109/chase.2016.14](https://doi.org/10.1109/chase.2016.14)]
31. Bleichert J, Liedlgruber M, Lender A, Reichenberger J, Wilhelm F. Unobtrusive electromyography-based eating detection in daily life: a new tool to address underreporting? *Appetite* 2017 Nov 01;118:168-173. [doi: [10.1016/j.appet.2017.08.008](https://doi.org/10.1016/j.appet.2017.08.008)] [Medline: [28797702](https://pubmed.ncbi.nlm.nih.gov/28797702/)]

Abbreviations

API: application programming interface

AUC: area under the curve

AUPRC: area under the precision–recall curve

Edited by G Eysenbach; submitted 13.02.21; peer-reviewed by M Manzanares, D Ravi, G Lim, OS Liang; comments to author 07.05.21; revised version received 28.05.21; accepted 30.12.21; published 01.03.22.

Please cite as:

Zhang B, Deng K, Shen J, Cai L, Ratitch B, Fu H, Guan Y

Enabling Eating Detection in a Free-living Environment: Integrative Engineering and Machine Learning Study

J Med Internet Res 2022;24(3):e27934

URL: <https://www.jmir.org/2022/3/e27934>

doi: [10.2196/27934](https://doi.org/10.2196/27934)

PMID: [35230244](https://pubmed.ncbi.nlm.nih.gov/35230244/)

©Bo Zhang, Kaiwen Deng, Jie Shen, Lingrui Cai, Bohdana Ratitch, Haoda Fu, Yuanfang Guan. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 01.03.2022. 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.

Original Paper

Tracking Subjective Sleep Quality and Mood With Mobile Sensing: Multiverse Study

Koen Niemeijer¹, MSc; Merijn Mestdagh¹, PhD; Peter Kuppens¹, PhD

Faculty of Psychology and Educational Sciences, Katholieke Universiteit Leuven, Leuven, Belgium

Corresponding Author:

Koen Niemeijer, MSc

Faculty of Psychology and Educational Sciences

Katholieke Universiteit Leuven

Tiensestraat 102

Post box 3717

Leuven, 3000

Belgium

Phone: 32 16372580

Email: koen.niemeijer@kuleuven.be

Abstract

Background: Sleep influences moods and mood disorders. Existing methods for tracking the quality of people's sleep are laborious and obtrusive. If a method were available that would allow effortless and unobtrusive tracking of sleep quality, it would mark a significant step toward obtaining sleep data for research and clinical applications.

Objective: Our goal was to evaluate the potential of mobile sensing data to obtain information about a person's sleep quality. For this purpose, we investigated to what extent various automatically gathered mobile sensing features are capable of predicting (1) subjective sleep quality (SSQ), (2) negative affect (NA), and (3) depression; these variables are associated with objective sleep quality. Through a multiverse analysis, we examined how the predictive quality varied as a function of the selected sensor, the extracted feature, various preprocessing options, and the statistical prediction model.

Methods: We used data from a 2-week trial where we collected mobile sensing and experience sampling data from an initial sample of 60 participants. After data cleaning and removing participants with poor compliance, we retained 50 participants. Mobile sensing data involved the accelerometer, charging status, light sensor, physical activity, screen activity, and Wi-Fi status. Instructions were given to participants to keep their smartphone charged and connected to Wi-Fi at night. We constructed 1 model for every combination of multiverse parameters to evaluate their effects on each of the outcome variables. We evaluated the statistical models by applying them to training, validation, and test sets to prevent overfitting.

Results: Most models (on either of the outcome variables) were not informative on the validation set (ie, predicted $R^2 \leq 0$). However, our best models achieved R^2 values of 0.658, 0.779, and 0.074 for SSQ, NA, and depression, respectively on the training set and R^2 values of 0.348, 0.103, and 0.025, respectively on the test set.

Conclusions: The approach demonstrated in this paper has shown that different choices (eg, preprocessing choices, various statistical models, different features) lead to vastly different results that are bad and relatively good as well. Nevertheless, there were some promising results, particularly for SSQ, which warrant further research on this topic.

(*J Med Internet Res* 2022;24(3):e25643) doi:[10.2196/25643](https://doi.org/10.2196/25643)

KEYWORDS

mobile sensing; sleep; subjective sleep quality; negative affect; depression; multiverse; multilevel modeling; machine learning; mood; mood disorder; mobile sensors; sleep quality; clinical applications

Introduction

Background

Sleep is important for well-being and plays a key role in people's mood and their risk for mental health problems such as depression. Therefore, obtaining a reliable picture of people's sleep quality critical for the study, prevention, and treatment of mood disorders. However, most of the existing methods are based on obtrusive techniques, either asking participants to wear dedicated technological devices (in the laboratory or otherwise) or to self-report their sleep patterns in a sleep diary. Therefore, both methods are burdensome for individuals and are not appropriate for long-term monitoring. If a method that is capable of unobtrusively tracking a person's sleep quality is developed, it would mean an important leap forward. One potential solution is the use of smartphones that have become ubiquitous in today's society. Approximately 76% of adults in countries with advanced economies have a smartphone [1], a device equipped with several sensors that automatically record user behaviors, some of which may be considered sleep-related ones. This study aims to examine to what extent automatically recorded information with smartphones—known as mobile sensing—can be used to predict self-reported subjective sleep quality (SSQ), mood, and the risk of mood disorders in the form of depression.

Sleep quality can be measured either in terms of subjective, self-reported sleep quality or can be captured with objective measurements such as polysomnography (PSG) and more recently actigraphy. Subjective and objective sleep quality are only weakly related [2-4] because they measure different aspects of sleep quality, namely physical sleep quality (objective) or a person's perception of sleep quality (subjective). However, they are both predictive of moods [5-7] and mood disorders [8-10]. Given the obtrusiveness of objective measurement methods and their adversity to use over long periods, we focus on predicting SSQ. Therefore, the primary outcome of this study is to measure SSQ with mobile sensing. As a secondary outcome, given the relevance of sleep quality for moods and mood disorders, we will examine to what extent mobile sensing is capable of predicting moods in terms of the daily negative affect (NA) and risk for mood disorders in terms of depression symptom severity. Thus, the underlying mechanism (and the underlying data on which this mechanism relies) is the relation between sleep and mental health.

Related Work

In most studies, an accelerometer is used to detect smartphone movements that are seen as synonymous with participant movements in bed [11-14]. Such movement detection is also used in commercial sleeping apps, such as Sleep as Android [15], Sleep Cycle [16], and Apple's native health app. Accelerometers can be used for movement detection with and without instructions to the participants. For example, researchers can instruct participants to place their smartphone next to or under their pillows. However, for the sake of ecological validity, this is usually not recommended. Another seemingly straightforward feature indicating sleep is whether the smartphone is being used (ie, screen on/off events) [14,17]. These screen events have proved useful for detecting circadian

rhythms and people's temporal sleep preferences [18]. A more contextual feature comes from the light sensor; ambient light indicates whether a room is dark and thereby helps determine whether a participant is sleeping [12,19,20]. Similarly, a microphone can pick up ambient sound, thereby confirming whether a participant is in a quiet environment [11,14,20].

Early efforts were made, for example, using location (GPS), contextual (sound and light), physical activity (accelerometer), and communication data (text and call logs) to infer SSQ measured with the Pittsburgh Sleep Quality Index [21], divided into 4 categories [22]. Using these features in a factor graph model, an accuracy of 78% was achieved. Other pioneering research on mobile sensing and sleep was conducted by comparing the quality of a smartphone accelerometer with an actigraph accelerometer for sleep monitoring [23]. Both devices showed reasonable agreement with all features except sleep onset latency. However, on a more critical note, 1 study [24] found no significant correlation between sleep measured by PSG and a smartphone app, although the mobile app did have high sleep-wake detection performance (85.9%) in relation to PSG. Thus, although mobile sensing for sleep detection has attained modest success, it is unclear how robust these methods are to changes in features and how well they describe sleep in comparison to PSG.

Despite PSG being considered the gold standard for sleep research, it is being recently rivaled by numerous wearable devices that feature an integrated accelerometer (among other sensors). Moreover, a mobile app was developed that achieved better user experience and lower perceived intrusiveness than wearables [11]. Using light, phone usage (including screen state), physical activity, and sound features in a completely "hands-off" approach, the app had an estimated sleep duration error of approximately 42 minutes, a slightly worse result than the other sleep detection methods, with 23 minutes when the smartphone is on the bed and 10 minutes for wearables. Subsequently, an app called StudentLife was developed with which a significant negative correlation among sleep, depression, and stress was found [19]. By using the same features as those in an earlier study [11], 95% of bedtimes were predicted with an error of 25 minutes from the ground truth as measured with an actigraph (Jawbone UP). Bringing the sensing accuracy of smartphones on par with that of wearables is significant because wearables are still not pervasive, thus making smartphones the least obtrusive solution for now.

A more indirect way of assessing sleep is by dividing a night into small epochs and predicting whether a participant is asleep or awake through self-reported bed and wake-up times or actigraphy. Of course, these sleep-wake states can then be used to derive other features that may subsequently be used for predicting sleep proxy measures. For example, 10-minute segments of sleep-wake states were predicted with 93% accuracy, but daily sleep quality was predicted with 84% accuracy [12]. Similarly, 15-minute bins were predicted with 89% accuracy using only screen on/off events [25]. Furthermore, 88.8% accuracy was reported when determining whether a 10-minute segment was a sleep state, whereas using the time of day alone achieved 86.9% accuracy [26]. Moreover, the accuracy per participant ranged from 65.1% to 97.3% although

lower accuracies could mostly be attributed to missing data reports and misreports. After adjustments, the accuracy was increased to 91.8%, corresponding to a median absolute deviation (MAD) of 38 minutes for the sleep start time and 36 minutes for the sleep end time.

This Study

We analyzed data from a 2-week study involving various smartphone sensors and the experience sampling method (ESM, also known as ecological momentary assessment) to collect SSQ in the morning and daily moods in terms of NA, as well as a 1-time assessment of person-level depression using a validated depression screening instrument. As mentioned above, several theoretically and empirically based predictions can be made regarding what sensors are the most significant for predicting SSQ and the associated variables of daily moods and depressive symptom severity. In addition, there are multiple ways to process the data based on different thresholds and data combinations, and there are different options to model the relationship between mobile sensing and the predicted variables. To examine this variability, we perform a multiverse analysis [27] in which we compare prediction results across multiple sensors, extracted features, thresholds, outlier methods, and prediction models. In other words, essentially all sensible combinations of these aspects, choices referred to as multiverse parameters, are evaluated in terms of how they are capable of predicting the outcome measures. Using this approach, this study not only aims to predict sleep with mobile sensing but also to investigate (1) which model, (2) which feature, and (3) which sensor works best within the context of the study. Consequently, this method is an analysis of robustness and transparency because it shows how different choices affect the outcome.

Methods

Overview of Data Preprocessing

As already mentioned earlier, we focused on the many seemingly arbitrary choices to be made during data preprocessing. Consequently, we performed a multiverse study showing the robustness of these choices. Following this approach, this section describes each aspect of the multiverse approach separately. First, various preprocessing steps were conducted for minimizing the number of participants with faulty data, namely removing participants with fewer than 10 data points for a sensor, sensors where all values were 0 (broken sensor), and 1 participant who had 40,000 times more data points than the participant with the second highest number of data points. In this regard, we present preprocessing of the different sensors in this study and describe when they are considered to indicate that a participant is sleeping, often having more than 1 threshold. Next, we discuss how features were built. These features are used in various statistical models. In the context of a multiverse approach, we experimented with several models to see which one worked best. Finally, we tried different outlier

removal methods to catch faulty data entry and unlikely data instances.

Data Collection

Data were collected during a 2-week study in 2018 [28]. Participants installed 2 separate apps on their smartphones, 1 for collecting sensor data and 1 for tracking their moods and SSQ using ESM. Mobile sensing data were gathered from specific (software) sensors to account for behavioral and contextual factors. These sensors are described in Table 1. Participants were instructed to charge their smartphones at night and connect to Wi-Fi whenever possible. On the other hand, data on the participants' SSQ and NA were collected using ESM, where they were prompted 10 times per day with a 16-item questionnaire measuring, among other parameters, their momentary (positive or negative) affect and how they slept if it was their first survey of the day. Depression levels were recorded only after the study using the Depression, Anxiety and Stress Scale questionnaire [29,30]. Katholieke Universiteit Leuven's Social and Societal Ethics Committee, whose directives are based on the Helsinki Declaration, approved the study (reference no. G-2018 01 1095). Written electronic consent from all subjects included in this study was obtained.

In total, 230 people responded to the participant selection questionnaire posted on social networking sites and other places frequented by students at the University of Leuven. Participants were excluded if they (1) did not understand Dutch, (2) did not have sufficient activity on their devices, (3) did not own an Android smartphone (except Huawei, Wiko, Medion, or Xiaomi), and (4) did not grow up with smartphones (ie, below 32 years of age). Among the 114 individuals who participated in an information session, 69 decided to join; 2 refused to sign the informed consent, and app installation failed for another 2 individuals. After excluding 5 participants who completed fewer than 30 ESM surveys, 60 participants remained. Furthermore, all data were preprocessed such that participants with fewer than 10 observations in total or only 0s (indicating a broken sensor) were left out. Single observations were left out if they were far away in time from any other observations (ie, if there was a gap in time in the data) using boxplot outliers. Next, values for each sensor were visually inspected to exclude participants who had very poor data quality. Participants were only retained if they had sufficient data for all sensors (ie, more than 10 observations), resulting in a final data set of 50 participants. Because participants were mainly university students and advertising was mainly done on social media and the campus of the faculty of psychology, the majority of the sample was young and female (14 male and 36 female aged 17-32 years, mean 21.90, SD 2.38 years).

Sensor Data Preprocessing

In this study, we used the following sensors: accelerometer, charging status, light sensor, Android activity recognition (AAR), screen activity, and Wi-Fi status. Table 1 gives an overview of the sensors and how their data were handled to obtain information on participants' sleep quality.

Table 1. Description of sensors used in this study^a.

Sensor	Description	Considerations	Indication of sleeping when...
Accelerometer	Acceleration in m/s ² along the x-axis	Other axes (y and z) not measured owing to programming error	Absolute value is less than (0.25, 0.5, 1) m/s ² .
Charging	Indication of whether the smartphone is charging As participants were instructed to leave their smartphone charging at night, a charging smartphone can mean that the participant is sleeping.	Momentary sleep interruptions are unlikely to be picked up by this sensor.	Battery is charging.
Light	Illuminance in lux ^b captured by the light sensor (usually at the front-top side of the smartphone)	Room with translucent curtains may simulate a wake-up state. Smartphone may be upside down.	Value is less than (8, 12, 16) lux.
Physical activity	Category of physical activity as measured by AAR ^c (ActivityRecognitionApi app)	Proprietary algorithm obfuscates how AAR is transformed into discrete activities.	AAR returns "Still."
Screen activity	Logs whether the screen has been changed to on or off.	Duplicate entries (screen off multiple times without screen on) are frequent. There is no distinction regarding whether the app turned the screen on (eg, notifications) or the participants.	Screen is off.
Wi-Fi	Information about whether the smartphone is connected to Wi-Fi	Unlikely to pick up temporary wake-up states	Connected to Wi-Fi.

^a The considerations column discusses issues that are (likely) associated with the sensors or their implementation. The sleeping when... column indicates when a sensor decides that a participant is sleeping in a window. Multiple values in parentheses indicate that the same data have been used but with a different cutoff value.

^b lux is the SI unit for intensity.

^c AAR: Android activity recognition.

One of the core elements in this multiverse study is inspecting each sensor and applying suitable preprocessing. The accelerometer was thought to be highly indicative of the sleep state [14,20]. However, as this sensor was sampled only every 2.9 minutes on average (SD 6.72 minutes with the longest interval being 15 minutes), only movement at that specific moment leads to the detection of a nonsleeping state. Additional problems with this sensor were that owing to a programming error, only the x-axis data were collected and the sensor in the participants' smartphones was uncalibrated. The latter issue led to some accelerometers not showing 0 acceleration at their lowest point. To offset this, accelerometer values were centered per participant, and the absolute values were calculated subsequently, such that a corresponding sleep or wake state could be derived from the threshold values specified in Table 1. As the accelerometers were not calibrated beforehand, we attempted to align them by centering. Moreover, for every acceleration, there must be an opposite deceleration. Physical activity is the product of on-the-fly processing by Google's AAR [31]. In brief, AAR runs an algorithm that combines multiple sensors to form discrete activities. In this research, only the activity "Still" was used because this indicates that the smartphone is not moving (whereas the other activities indicate some type of movement).

The screen state software sensor was preprocessed in such a way that interactions of less than 12 seconds were ignored. This is necessary because some of the smartphone screens light up whenever there is a new notification and thus pollute the actual

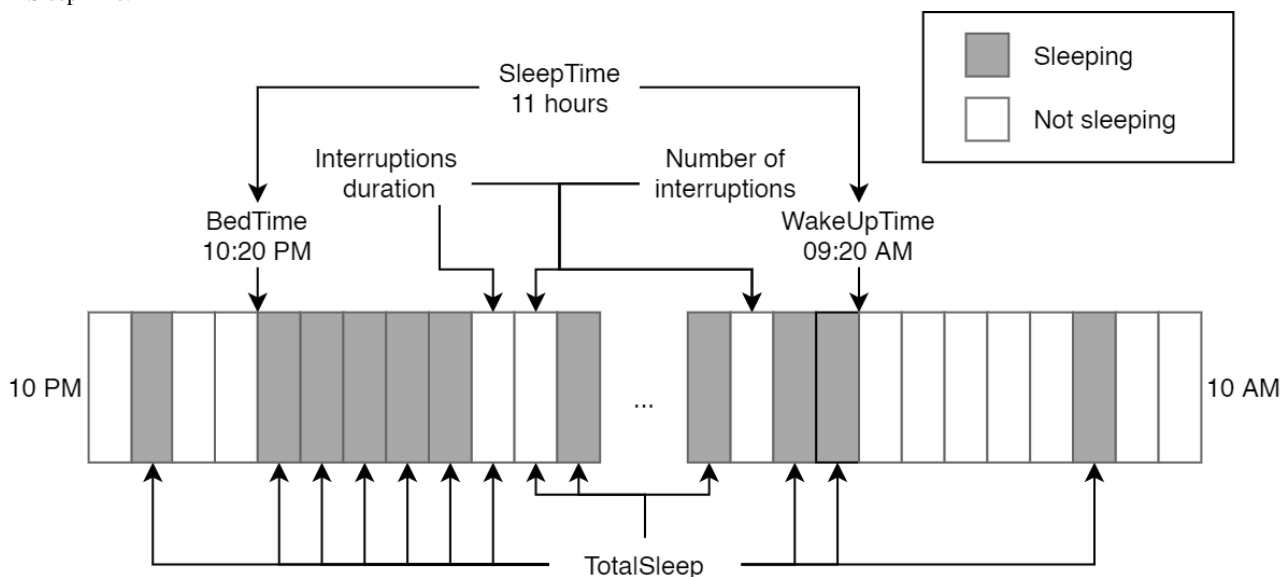
stream of the screen state. On the other hand, if people use their smartphone in the middle of the night for, say, checking the time, this will most likely also be thrown out. For the charging, light, and Wi-Fi sensors, no further preprocessing was applied.

After cleaning the data, the participant's nights were divided into windows of 5 minutes. To avoid a high number of false positives, only data from 10 PM to 10 AM on the next day were considered. For each window and sensor, a majority vote determined whether a participant was sleeping. For example, if the app sensed that the smartphone was charging twice and that it was not charging once (within the same window), a majority vote set the window to sleeping. If there were no data in the window, the last observations (ie, of the previous window) were carried forward.

Feature Creation

Several features were extracted from the data that were deemed indicative of sleep quality. Figure 1 displays a visual example of all the features explained in this section. The basic building blocks for most features are the BedTime and WakeUpTime features that are found by examining 3 to 5 adjacent windows where a participant is sleeping (BedTime) or not sleeping (WakeUpTime) according to a sensor. Concretely, for the first 3 to 5 adjacent windows where a participant is sleeping, the first window of that sequence is the BedTime window. The opposite is true for WakeUpTime, namely the first window of the first 3 to 5 adjacent windows where a participant is not sleeping.

Figure 1. A fictional example of the features used in this study. This figure only depicts the first and last hours of the windows for illustration purposes, where the ... obscures the rest of the night. Gray bars indicate sleep states, and white ones indicate nonsleeping states. The night starts at 10:20 PM (ie, before the first sleeping window) because BedTime requires 3 to 5 adjacent windows (4 windows in this example). Similarly, WakeUpTime starts at 9:20 AM because this denotes the start of 3 to 5 adjacent windows, depending on the chosen parameters. Notice how SleepTime is the time between BedTime and WakeUpTime, whereas TotalSleep is simply the sum of all the sleeping windows. Finally, the number of interruptions is given as the number of nonsleeping sequences minus the number of sleeping sequences; the interruptions duration feature includes all the nonsleeping windows within SleepTime.



From the initial features of BedTime and WakeUpTime, several others could be calculated as well. First, SleepTime was calculated as the time between BedTime and WakeUpTime. Second, sleep quality may be impacted when a participant sleeps unusually longer or shorter than usual. To capture this, we computed the deviation from each participant’s average BedTime and WakeUpTime (in hours) for each night (abbreviated as devAvgBedTime and devAvgWakeUpTime, respectively). Hours have been recentred such that 10 PM=0, 11=1..., 10=12. Third, the number and duration of nightly interruptions (abbreviated as nInterruptions and InterruptionsDuration, respectively) are calculated as follows: Every pattern of neighboring wake-up minus the sleeping windows counts as one interruption, and the total number of nonsleeping windows forms the total duration (multiplied by the window size of 5 minutes). Both features are only calculated within the interval of BedTime – WakeUpTime, also known as the SleepTime.

As we did not know the validity of BedTime and WakeUpTime—based on which all features discussed so far—we added a feature called TotalSleep that is simply the number of sleeping windows at night (10 PM-10 AM) multiplied by the window size (ie, 5 minutes). Another feature not based on sleep or wake-up times but can be useful is the physical activity of the previous day, as measured by AAR. Concretely, UserActive is the sum of the time spent walking or cycling during the previous day.

Modeling Approach

As the purpose of this research was to predict SSQ, which cannot be measured directly, we used proxy measures to validate our approach. In particular, we trained models to predict the SSQ reported each morning, average NA of the next day, and the participants’ general level of depressive symptoms. All these

variables are on a scale of 0 to 100. The input features used were those described in the previous section (ie, SleepTime, devAvgBedTime, devAvgWakeUpTime, nInterruptions, InterruptionsDuration, UserActive, and TotalSleep). For better understanding the effect of independent variables on dependent variables, only 1 input feature (eg, SleepTime) per dependent variable was used. Therefore, 21 combinations of independent and dependent variables had to be tested. We also attempted using several models to see which one worked best. We applied multilevel models (MLMs) [32], k-nearest neighbors (kNN) [33], support vector machines (SVMs) with a radial kernel [34], extreme gradient boosting (XGB) [35], and generalized additive models (GAMs) with smoothing splines [36]. For predicting depression, linear models (LMs) were used instead of MLMs because these data do not have a nested structure (ie, every participant has only 1 data instance). All these models were trained and tuned using the caret package in the R statistical software package [37]. Because there were several missing sleep values (see the Results section), we decided to apply only generalized and not personalized models because these would be left with too few observations.

Outlier Analysis

Because of the noisy nature of this type of data, outlier analysis may be needed to remove unlikely observations. Several strategies are available to detect outliers, including a strategy where no outliers are removed. A simplistic method for doing this is by subtracting a multiple of the SD from the median. First, the median was chosen because some variable distributions may be significantly skewed, and secondly, we chose to multiply the SD by 3 because we only wanted to throw out those observations that were truly outliers. A slightly adjusted version of this method is one that does not use the SD but uses the MAD because this is often a more robust measure of dispersion.

Finally, a multivariate outlier method was also applied, namely isolation forests that work by building a decision tree to isolate anomalies [38,39]. Outlier observational strategies were also parameters evaluated in this multiverse analysis and are given in Table 2.

Multiverse

As mentioned earlier, an important aim of this study is to evaluate the effect of several data preprocessing, sensor, and modeling choices on the prediction of sleep quality, mood, and depression via a multiverse study. Table 2 gives an overview of all the multiverse variables.

Table 2. Multiverse parameters and values considered in this study^a.

Parameter	Number of values (N)
Sensors	10
Sensor thresholds for sensors listed in Table 1	
Features	7
devAvgBedTime	
devAvgWakeUpTime	
InterruptionsDuration	
nInterruptions	
SleepTime	
TotalSleep	
UserActive	
Outcome variables	3
Subjective sleep quality	
Negative affect	
Depression	
Models	5
Generalized additive models with smoothing splines	
k-nearest neighbors	
Mixed effects models	
Support vector machines	
Extreme gradient boosting	
Outlier removal	4
None	
median(x) ± 3σ _x per participant	
median(x) ± 3 × mad ^b (x) per participant	
Isolation forests; Liu et al [38]; Cortes [39]	
Feature thresholds	3
3 to 5 adjacent windows	
Total number of combinations	12,600

^aFor each parameter, N indicates how many values are there so that the total number of combinations is the product of N.

^bmad: median absolute deviation.

Because there are 12,600 combinations in this multiverse setup, overfitting plays an even larger role than usual; that is, simply splitting the data set into a training and test set would lead us to choose the model that overfits the training and test sets the most. Therefore, our workflow to minimize overfitting followed a hold-out set approach given below [40]:

1. Split the data into training, validation, and test sets.
2. Train all models on the training set.

3. Validate these models on the validation set based on R^2 .
4. For each outcome variable, choose models (maximum of 5) within 1 SE from the best model (known as the 1-SE rule) [41,42].
5. Test these best models on the test set using R^2 .

In this procedure, the training set contains the data of the first week whereas the validation and test sets contain data of days

1 to 4 and days 5 to 7 of the second week, respectively. Alternatively, as we used only 1 data instance per participant when predicting depression, the data set was split such that 60%, 20%, and 20% of the participant data used for the training, validation, and test sets, respectively, were stratified by gender. By doing this, we extracted only the best models based on the validation set and then used them on the test set to confirm if they overfitted the test set. Results were measured by their predicted R^2 [43], a performance metric that is calculated as follows:



where n represents the number of data points, Y is the vector of the observed values, and \hat{Y} is the vector of the predicted values. In essence, this representation of R^2 is a scaled version of the commonly used mean squared error when applying machine learning models or cross-validation.

Results

Missing Data

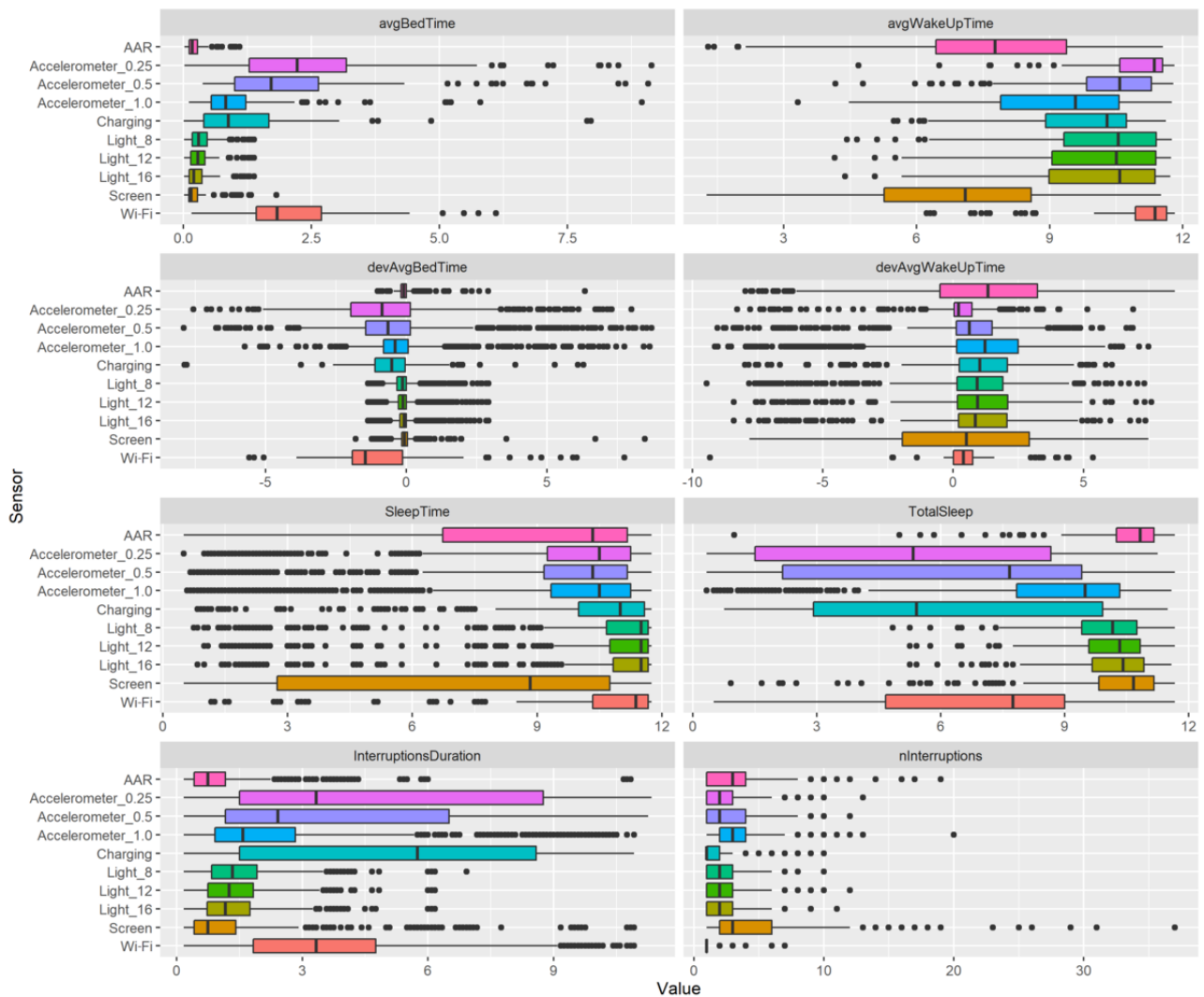
An important consideration in this type of data is missingness. There are multiple reasons for this. First, participants were

prompted with only 14 beeps asking about SSQ in total, so missing 1 beep could have led to the loss of valuable data. Second, preprocessing and outlier removal techniques are responsible for deleting even more data. This coincides with a third major reason for missing data; the wake-up time was often not found for several sensors. This is especially true for the light sensor possibly because it could not detect enough instances where the illuminance was higher than the threshold because the threshold was too high, the phone was upside down, or the phone disappeared inside the participant's pocket quickly after waking up; there could be other reasons as well. In our analyses, missing data were not imputed but considered as missing at random.

Descriptive Statistics

To obtain a better understanding of what the features encompass, Figure 2 presents boxplots for each feature per sensor. The displayed variation can be interpreted as how the values of the features change based on what settings (excluding the prediction model choices) are chosen for the multiverse parameters. Therefore, this variability is not related to the variability of features over participants; it purely highlights the robustness of the features with regard to the multiverse choices.

Figure 2. Boxplots showing how different settings in this multiverse study (for parameters see Table 2) affect each feature per sensor. The smaller the range of the boxplot, the more robust a feature is to multiverse choices. The feature for physical activity performed the previous day (UserActive) is not shown because this value does not vary with any multiverse parameter except in outlier analysis. Observe that the x-axis shows hours except in the case of nInterruptions, which shows a count. Moreover, avgBedTime and avgWakeUpTime are influenced by the criterion for defining a night, namely from 10 PM to 10 AM.



In general, the values within the features in Figure 2 vary considerably, depending on the sensors and other multiverse parameters. There are many outliers that can either represent true patterns or that occur because of data anomalies. Because of the explorative nature of this study and the many combinations we attempted, we assumed the actual patterns to be restricted to the boxes and whiskers, with outliers representing rare behavior or, more likely, data errors. We will now describe some of the key observations and patterns regarding specific features.

First, notice that devAvgBedTime and devAvgWakeUpTime are skewed toward 0. This occurs because these features are the deviations from the mean bedtime or wake-up time. Deviations for bedtime are usually not too high, but for the wake-up time, they seem more problematic. For bedtime, the range for some sensors is very small and skewed toward 0 (ie, at 10 PM), indicating that on average, participants went to bed as soon as the night started or possibly even earlier, no matter what values the multiverse parameters have. On the other hand, the spread for wake-up time is rather large, indicating that changing

multiverse parameters substantially impacts this feature. A second observation is that SleepTime appears to be more stable than TotalSleep, although this is not true for every sensor (eg, screen state).

One of the aims of conducting this multiverse experiment was to consider different thresholds for the light sensor and accelerometer. The light sensor provides consistent results for most features (ie, less spread), but this is mostly because almost all participants' bedtime is immediately at 10 PM. In other words, the highest variability appears to stem from a fluctuating wake-up time. For the accelerometer, a lower threshold corresponds to a later bedtime, a later wake-up time, and longer nightly interruptions. Furthermore, SleepTime is similar between different light and accelerometer thresholds, whereas TotalSleep is vastly different for the accelerometer but not for light.

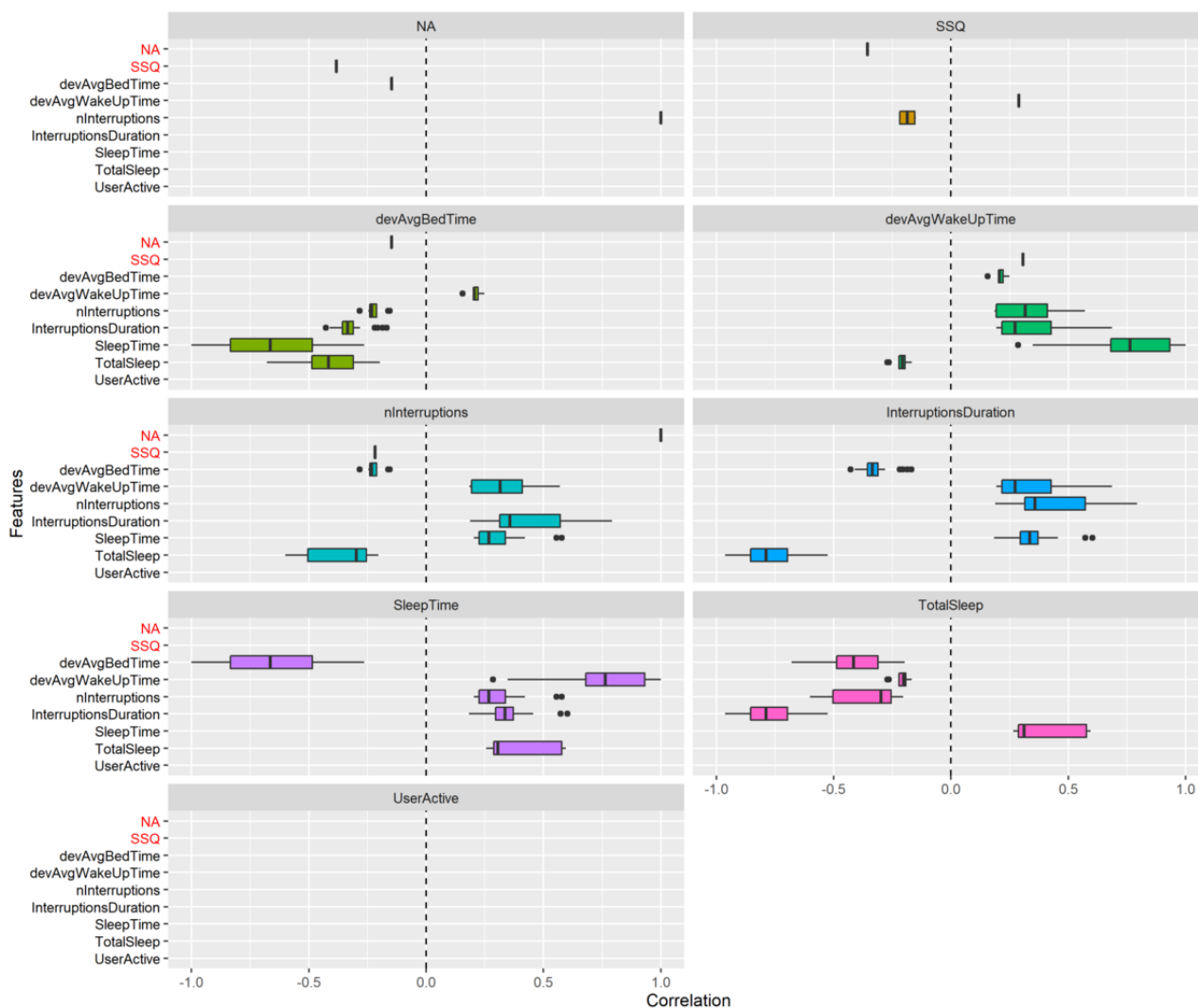
Correlations

To better understand how different features relate to each other, Figure 3 shows significant correlations between features, between outcome variables, and between features and outcome

variables. Correlations were extracted from mixed effects models as the coefficients of fixed effects using standardized inputs. The *P* values were calculated using analysis of variance. All correlations have been corrected for multiple tests [44]. Figure 3 shows that significant correlations between variables are scarce; for instance, UserActive is not related to any other

variable and the outcome variables NA and SSQ are also sparingly related to other variables. In fact, NA is only related to SSQ and the number of interruptions (nInterruptions) whereas SSQ is also only related to NA, nInterruptions, and the deviation from the average wake-up time (devAvgWakeUpTime).

Figure 3. Distribution of significant correlations between variables on the y-axis and in the facets. Note that in contrast to Figure 2, this plot does not show AvgBedTime and AvgWakeUpTime because there is only a single value per participant. SSQ and NA (highlighted in red) are outcome variables denoting the subjective sleep quality and negative affect, respectively. The variance in the boxplots represents how the correlation changes with different values for the multiverse parameters (including different sensors). Without the boxplot, there was no significant correlation. This plot is symmetrical, meaning that every relationship is also shown oppositely. Note that when the boxplot is reduced to a single bar, this likely means there is only a single value (or less likely that there are multiple values very close together).



For depression, correlations were not extracted from an MLM (and hence not shown in Figure 3) but rather as Pearson correlations because there is only a single observation per participant. For all variables, the mean was used to generate correlations with depression scores, and multiple testing correction was subsequently applied [44]. There was 1 significant negative correlation with depression and the feature SleepTime ($r=-0.46$) using the accelerometer (1 m/s^2), and 7 negative correlations with the outcome variable SSQ ($r=-0.46$ to -0.51) (0.25 to 1 m/s^2). Furthermore, there was 1 significant positive correlation between depression and NA ($r=0.51$) using AAR.

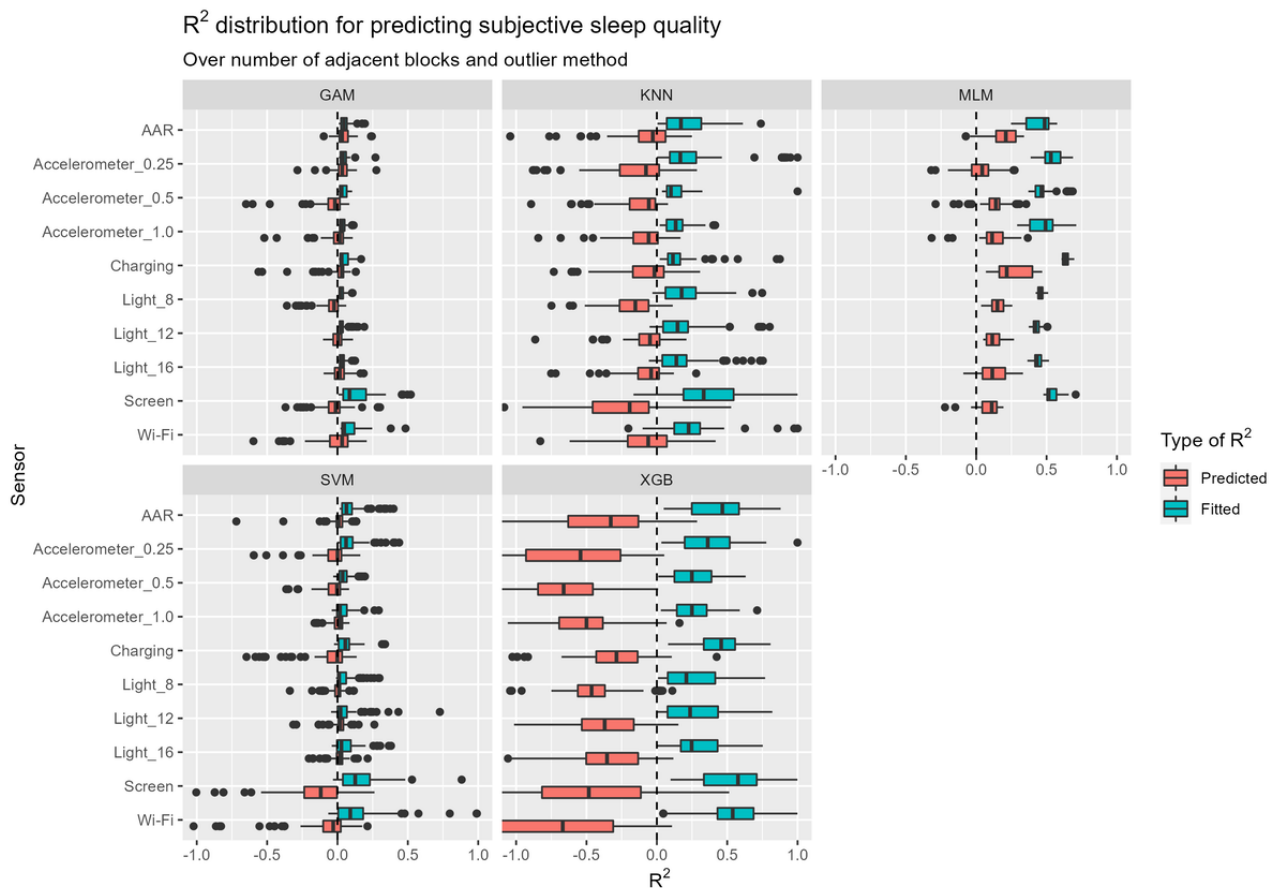
Evaluation

As described previously, the data set was divided into training, validation, and test sets. To gain a better understanding of how each sensor and sleep feature relates to the output variables, we trained separate models for each feature, sensor, and model combination. By doing so, we could inspect their performance (measured in terms of R^2) on the validation set in relation to the other multiverse parameters. After doing so, we selected the best (maximum 5) models within 1 SE from the best model (ie, the model with the highest R^2 on the validation set) and tested them on the test set to draw definitive conclusions.

First, we present the results of our predictions of SSQ on the validation set in Figure 4, where we can see the predictive performance of different models and the impact of selecting a specific sensor varying with other multiverse parameters (as specified in Table 2). In this plot, it is clear that the MLMs perform the best and the XGB models perform the worst in terms of the predicted R^2 on the validation set. Moreover, the GAMs and SVMs perform similarly, having an R^2 of approximately 0 with extreme values on either side. The kNN models show a slightly poorer performance, usually with an R^2 below 0. In terms of more general patterns, we can see that the models have generally overfitted the data; that is, the models have achieved a higher R^2 on the training set than on the validation set. However, the boxes and whiskers of the boxplots represent only the majority of the results, and even a single

outlier may represent a model that does not overfit and performs well. After all, the best models that we selected to be tested on the test set are probably “outliers” in these figures. As far as sensors are concerned, the charging sensor seems to perform slightly better, at least when using MLMs or XGB models but not for others. The other sensors appear to be approximately equivalent in performance. As observed in Multimedia Appendix 1, we replaced the sensors on the y-axis with features so that we could obtain an idea of how different features affect the outcome. As these boxplots are also separated based on the models used, they have approximately the same R^2 distribution, as shown in Figure 4. The performances of features are relatively similar. One might argue that nInterruptions generally performs slightly better in combination with XGB models and MLMs, but this effect is small.

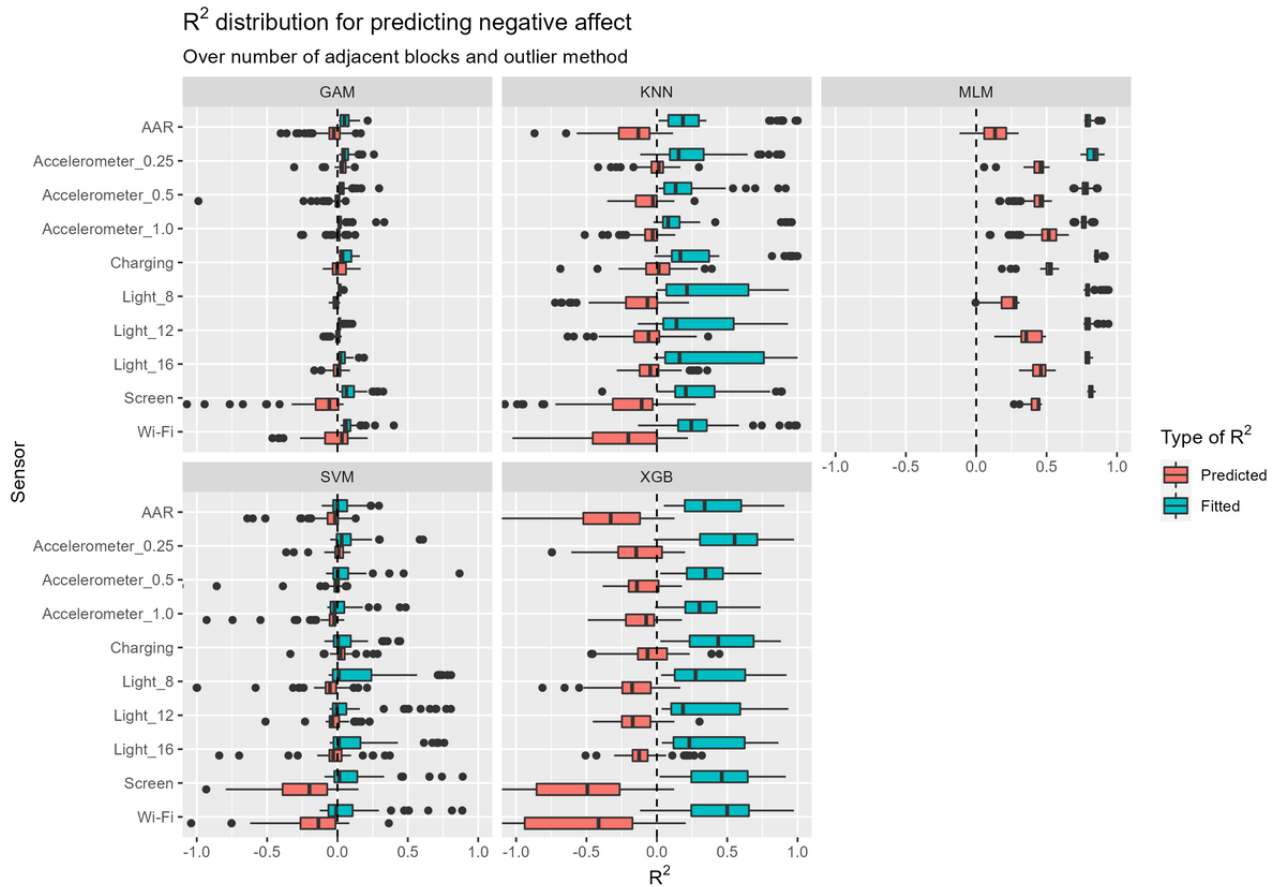
Figure 4. "R2" distribution for predicting subjective sleep quality. The spread of the boxplots represents how the "R2" of predictions of subjective sleep quality change with varying multiverse parameters, namely different sensors, features, models, number of adjacent windows, and outlier methods. GAM: generalized additive model; kNN: k-nearest neighbors; MLM: multilevel model; SVM: support vector machine; XGB: extreme gradient boosting.



Second, the predictive performance regarding NA on the validation set is represented visually in Figure 5. Again, for most models, the MLMs appear to perform best as they show the highest R^2 on the validation set. Nevertheless, as indicated by the high discrepancy in R^2 values between the training and validation set, models are once again overfitting. Similar to that

observed in the prediction of SSQ, the XGB models perform the worst followed by the kNN models, and finally GAMs and SVMs, showing a predictive R^2 value of approximately 0. A notable difference among these models is that there is little variation among the MLMs, GAMs, and SVMs, meaning that multiverse parameter choices have less impact than that observed for the kNN and XGB models.

Figure 5. "R²" distribution for predicting negative affect. The spread of the boxplots represents how the "R²" of predictions of negative affect change with varying multiverse parameters, namely different sensors, features, models, number of adjacent windows, and outlier methods. GAM: generalized additive model; kNN: k-nearest neighbors; MLM: multilevel model; SVM: support vector machine; XGB: extreme gradient boosting.

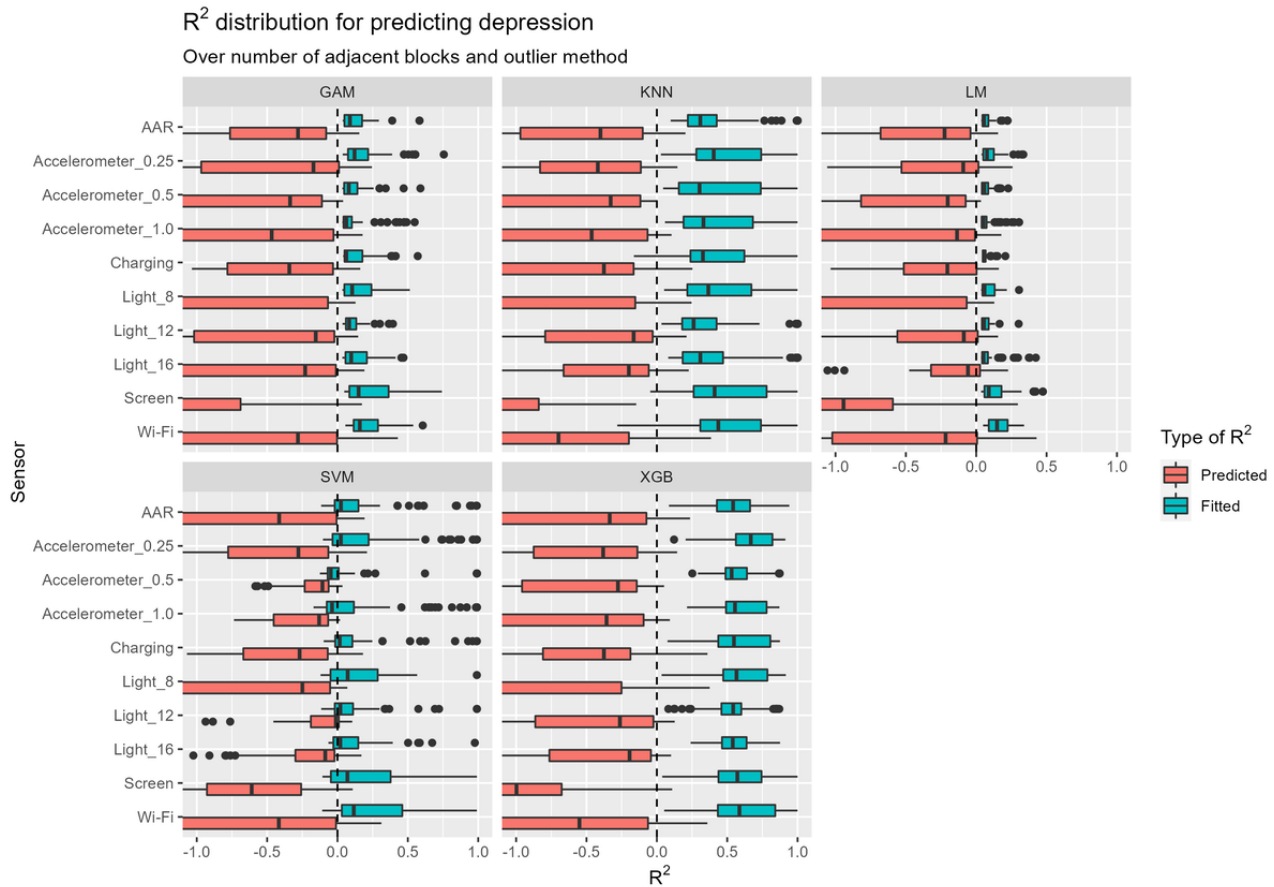


When predicting NA, choosing a specific sensor does not seem to matter, except for screen activity and Wi-Fi because these performed much worse. Indeed, all MLMs using the Wi-Fi sensor were found to be inoperable, indicating a serious problem with these models (eg, not enough data, no convergence). Compared to the rest of the sensors, screen activity and Wi-Fi also performed much worse when using XGB models. When finding the best performing feature, [Multimedia Appendix 2](#) shows a trend similar to that of the sensors, meaning that there is no single feature that is better than all the rest.

Finally, [Figure 6](#) shows the performance distribution when predicting depression in relation to other multiverse parameters.

Contrary to SSQ and NA, the model performance spread in terms of R^2 for depression is much greater, implying that the selection of the multiverse parameters influences model performance. There are considerable differences among sensors too. For example, charging and AAR appear to perform better across all models than the other sensors; that is, because their spread is smaller, models using the charging or AAR sensor are likely to perform better. On the other hand, the 1 m/s² accelerometer has a much larger spread and a lower mean R^2 than the other sensors.

Figure 6. "R²" distribution for predicting depression. The spread of the boxplots represents how "R²" values when predicting depression change with varying multiverse parameters, namely different sensors, features, models, number of adjacent windows, and outlier methods. GAM: generalized additive model; kNN: k-nearest neighbors; MLM: multilevel model; SVM: support vector machine; XGB: extreme gradient boosting.



Multimedia Appendix 3 presents similar findings about the features when predicting depression. That is, the predictions are much worse than those for other outcome variables. The nInterruptions feature appears to perform slightly better than the others although this applies only to the spread, implying that with this feature, models perform better on average but that there may be models from other features performing much better.

Although it is important to describe how different sensors and features impact the outcome variables on average, our ultimate objective is to find how we can best predict the outcome variables from the available data. To this end, Figure 7 presents the results of the best models per outcome variable.

Unsurprisingly, the sensors and features of these best models are generally also the ones that performed the best on average for those outcome variables. However, there is a remarkably large gap between the R^2 values on the test set and those on the training and validation sets. For example, a kNN (k=7) model for predicting SSQ using the screen state sensor, 5 adjacent windows, isolation forests, and the feature devAvgWakeUpTime achieved R^2 values of 0.558 and 0.528 on the training and validation sets, respectively. As this does not show much overfitting, one would expect the test R^2 to be similar. However, the R^2 on the test set is only -0.531, revealing the model to be less than informative.

Figure 7. "R2" distribution for predicting depression. The spread of the boxplots represents how "R2" values when predicting depression change with varying multiverse parameters, namely different sensors, features, models, number of adjacent windows, and outlier methods. GAM: generalized additive model; kNN: k-nearest neighbors; MAD: median absolute deviation; MLM: multilevel model; SVM: support vector machine; XGB: extreme gradient boosting.

	Sensor	Model	# of Windows	Outlier Method	Feature	Training R^2	Validation R^2	Validation SE	Test R^2
Subjective Sleep Quality	Charging	MLM	3	Isolation forests	nInterruptions	0.658	0.469	0.102	0.348
	Charging	MLM	3	Isolation forests	devAvgWakeUpTime	0.619	0.461	0.106	0.334
	Charging	MLM	3	Isolation forests	SleepTime	0.624	0.450	0.107	0.303
	Screen state	XGB	5	MAD	devAvgWakeUpTime	0.501	0.514	0.238	-0.003
	Screen state	KNN	5	Isolation forests	devAvgWakeUpTime	0.558	0.528	0.167	-0.531
Negative Affect	Accelerometer_1.0	MLM	5	MAD	InterruptionsDuration	0.779	0.62	0.101	0.103
	Accelerometer_1.0	MLM	4	Isolation forests	UserActive	0.767	0.624	0.098	0.081
	Accelerometer_1.0	MLM	5	MAD	devAvgBedTime	0.828	0.623	0.107	0.074
	Accelerometer_1.0	MLM	4	Isolation forests	InterruptionsDuration	0.759	0.622	0.099	0.066
	Accelerometer_1.0	MLM	5	MAD	UserActive	0.831	0.659	0.073	0.005
Depression	Wi-Fi	GAM	4	MAD	SleepTime	0.074	0.429	0.126	0.025
	Wi-Fi	LM	4	MAD	SleepTime	0.074	0.429	0.126	0.025
	Wi-Fi	KNN	3	None	UserActive	0.477	0.385	0.230	-0.639
	Wi-Fi	KNN	3	SD	UserActive	0.477	0.385	0.230	-0.639
	Light_8	XGB	5	MAD	devAvgWakeUpTime	0.760	0.375	0.296	-0.688

Figure 7 shows that SSQ can be predicted reasonably well; MLMs using the charging sensor, 3 adjacent windows, and isolation forests for detecting outliers achieved a test R^2 of 0.303 to 0.348 with nInterruptions, devAvgWakeUpTime, or SleepTime. All these features were positively related to SSQ, having a slope of approximately 0.442–0.643 with a high 95% CI –3.06 to 4.35. However, predicting NA proved to be more challenging because we obtained a highest R^2 of only 0.103 on the test set, a value that is considerably lower than that obtained on the training and validation sets. Nevertheless, it is worth noting that all the best models for predicting NA are MLMs using the accelerometer with a threshold of 1 m/s². Of these models, the ones using InterruptionsDuration and devAvgBedTime were positively related to NA (slopes of 0.0449–0.279 and 0.427, respectively) whereas userActive was negatively related (slope of 0.199–1.37). Finally, predicting depression was the most difficult task because only 1 best model yielded a barely informative result, namely the LM with an R^2 of 0.025. The predictions for SSQ appear better because it is more closely associated with objective sleep than the other outcome variables, namely NA and depression.

Discussion

Principal Findings

Although various studies have used different methods and sensors to successfully track sleep with mobile sensing [14,25,26], it is not yet clear how these various approaches impact the results. To this end, this study was designed to provide a transparent and robust method for tracking sleep with mobile sensing. By accounting for many choices in a multiverse study design, we were able to separate and elucidate the various effects of sensors, features, and models on sleep proxy measures, thereby providing an important step forward for future researchers. Principally—and in accordance with previous research—we showed that it is possible to explain a reasonable and an interesting proportion of the variance observed in SSQ (or the related variable of daily negative mood) using mobile sensing. Moreover, our multiverse analysis indicates which combinations of multiverse parameters (shown in Table 2) can do this best, at least in this sample. Although the results in this

study are slightly more pessimistic than those in other studies concerning mobile sensing and sleep (ie, some studies showing very high accuracies) [12,25,26], they cannot be compared directly given that we applied a strict cross-validation plan.

From the outlined results, we can conclude that for predicting SSQ, the charging sensor is overall the best performing and the most robust sensor although this sensor has a significant advantage because participants were instructed to leave their smartphone charging at night. However, for predicting NA, we found the accelerometer to be the best sensor despite its previously noted limitations (ie, low sample rate and measurements only along the x-axis). Based on previous research [11,14], this finding suggests that the accelerometer is a feasible option for sleep detection. Nevertheless, when predicting person-level depression, we found that the results varied considerably and were often less than informative. Even the best performing sensor achieved an R^2 of only 0.025 on the test set, and hence we could not draw any conclusions from this.

Another aspect we accounted for in this study is using multiple ways of constructing models (ie, we attempted multiple models to see which one worked best). The results point out that MLMs are clearly the best performing models for any outcome variable. However, it should be noted that the analysis is between participants whereas MLMs also use within-person information (although large amounts of within-participant data were missing). Likewise, LMs seem to be the primary choice for predicting depression as well, suggesting that robustness must be an important characteristic. Furthermore, GAMs and SVMs showed reasonably similar performances but much worse than that of MLMs. Finally, kNN and XGB models seemingly performed the worst, with XGBs displaying excessive overfitting.

Although not described in detail, Multimedia Appendix 1 presents the R^2 distributions of the results split by feature. We have already mentioned that no feature appears to be more important than any other. Another inference supported by Figure 2 is that some features are more robust than others in this multiverse analysis. Concretely, nInterruptions has less spread than the other features, indicating that it is a relatively robust

feature. This could be interesting for future studies in which researchers have to manually select features.

Limitations

There are a number of limitations that should be considered when interpreting the results of this study. The results of this study are generally limited by the relatively small size of the data set. This limitation poses a drawback because it makes the models more prone to overfitting, although this risk was mitigated by an elaborate hold-out set approach. Next, we list 3 distinct limitations that may have impacted the results.

First, we relied on self-reports from ESM as an estimate of sleep. This, combined with the fact that participants did not always answer the SSQ item in the morning, may have led to odd or spotty patterns. Despite trying to minimize the impact of this factor by predicting several outcome variables known to be associated with sleep (ie, not only SSQ but also NA and depression), we cannot be certain about the impact of this factor on the research results.

The second limitation resulted from several sensor issues during the trial. The most pressing issue is the low sample rate for the accelerometer, charging, light, and Wi-Fi, averaging approximately 1 measurement instance every 2.9 minutes. Especially for the accelerometer, this is not enough data because it only allows the sensors to record momentary changes instead of tracking participants' contextual surrounds for longer periods, as mentioned earlier. The accelerometer also did not collect the y- and z-axes data, which further limits its use. Furthermore, the way in which sensors collected data was directly influenced by participants' behavior in that they were instructed to keep the smartphone charged at night and be connected to Wi-Fi. This may mean the charging and Wi-Fi sensors are less ecologically valid.

Finally, we assumed that a normal night of sleep is somewhere between 10 PM and 10 AM the next day. Although this assumption is probably true for most people, students may not always follow a regular sleep schedule. Moreover, because some sensors are prone to produce many false positives (eg, the light sensor detecting the smartphone being in a participant's pocket as sleep), we felt it was necessary to reduce this at the cost of cutting off a small portion of the participants' sleep.

Contributions to Previous Work

Our work complements previous work described in various ways. First and foremost, this study developed and implemented a method to unravel the effects of distinct variables as well as sensors and their thresholds for the prediction of SSQ, NA, and depression. Such a method can guide future researchers intending to conduct similar studies to further examine these effects or help them choose which sensors to select for estimating sleep. It also provides a tool for researchers when they have to make choices based on the many different results that exist when predicting sleep with mobile sensing, such as the widely varying results in earlier studies [22-26]. Moreover,

it stresses the importance of making considerations and makes researchers aware that such choices may have a significant impact on their study, as suggested by the multiverse theory [27] and previous studies [26]. In addition, this study can help verify theories that link smartphone-recordable behaviors to sleep and mood, such as the purported links among sleep, depression, and stress [19].

Second, we showed that continuous SSQ and NA can be reasonably predicted from sleep features whereas depression cannot be predicted. Most studies reduce these features to a discrete variable with only 1 or 2 categories [12,22], whereas we took on the challenge to expand this to a continuous variable between 0 and 100. Naturally, a continuously scaled variable brings about more variation, but this variation can be approximated.

Future Directions

When performing such an explorative study where it is not yet clear how different variables relate to each other, it is important to at least have enough data to work with. Future studies should focus on obtaining a sufficiently large sample size to minimize the necessity of a potential multiverse study or at least minimize the risk of selecting a nonrepresentative sample. Additionally, missing data posed a big challenge in this study because participants did not respond to the first beep of the day.

Although this study has solely relied on SSQ and proxy measures such as NA and depression, in future studies, it could be useful to also collect self-reported bed- and wake-up times to validate these features. After all, several other features were based on these times; therefore, obtaining a better understanding of their accuracy also helps in further improving these features. Moreover, wearable devices are not as reliable as PSG, but they are considerably less obtrusive and have already been adopted by a sizable portion of the mainstream population to record sleep. As such, they can be added in future studies to achieve an additional high-quality estimate of sleep. Finally, if an approach could be developed to use mobile sensing data to estimate sleep (eg, through a multiverse approach), it should be compared to PSG as a final measure of validation.

Conclusions

Sleep plays an important part in moods and mood disorders but is difficult to track unobtrusively. Therefore, this study has shown that it is feasible to track sleep with mobile sensing although this depends strongly on which sensors, features, and models are chosen. In fact, our approach demonstrated that most combinations of multiverse parameters often form noninformative models. This can be partly attributed to issues with data collection and not having enough data for developing personalized models, but further research is needed to validate these problems. Moreover, SSQ predictions were better than the other proxy measures (ie, NA and depression severity). Nevertheless, the findings of this study are promising and warrant further investigation into the use of mobile sensing for tracking sleep.

Acknowledgments

The research that led to the findings reported in this article was supported by the research fund of Katholieke Universiteit Leuven (grants GOA/15/003 and C14/19/054).

Conflicts of Interest

None declared.

Multimedia Appendix 1

"R2" distribution for predicting subjective sleep quality. The spread of the boxplots represents how features change with varying multiverse parameters, namely different sensors, number of adjacent windows, and outlier method.

[[PNG File , 194 KB - jmir_v24i3e25643_app1.png](#)]

Multimedia Appendix 2

"R2" distribution for predicting negative affect. The spread of the boxplots represents how features change with varying multiverse parameters, namely different sensors, number of adjacent windows, and outlier method.

[[PNG File , 194 KB - jmir_v24i3e25643_app2.png](#)]

Multimedia Appendix 3

"R2" distribution for predicting depression. The spread of the boxplots represents how features change with varying multiverse parameters, namely different sensors, number of adjacent windows, and outlier method.

[[PNG File , 181 KB - jmir_v24i3e25643_app3.png](#)]

References

1. Taylor K, Silver L. Smartphone ownership is growing rapidly around the World, but not always equally. Pew Research Center. Washington DC, United States: Pew Research Center; 2019 Feb 05. URL: <https://www.pewresearch.org/global/wp-content/uploads/sites/2/2019/02/Pew-Research-Center-Global-Technology-Use-2018-2019-02-05.pdf> [accessed 2020-06-26]
2. Hernandez R, Vu THT, Kershaw KN, Boehm JK, Kubzansky LD, Carnethon M, et al. The association of optimism with sleep duration and quality: findings from the coronary artery risk and development in young adults (CARDIA) study. *Behav Med* 2020 Apr;46(2):100-111 [FREE Full text] [doi: [10.1080/08964289.2019.1575179](https://doi.org/10.1080/08964289.2019.1575179)] [Medline: [31339814](https://pubmed.ncbi.nlm.nih.gov/31339814/)]
3. Sutin AR, Gamaldo AA, Stephan Y, Strickhouser JE, Terracciano A. Personality traits and the subjective and objective experience of sleep. *Int J Behav Med* 2020 Aug;27(4):481-485 [FREE Full text] [doi: [10.1007/s12529-019-09828-w](https://doi.org/10.1007/s12529-019-09828-w)] [Medline: [31755033](https://pubmed.ncbi.nlm.nih.gov/31755033/)]
4. Unruh ML, Redline S, An MW, Buysse DJ, Nieto FJ, Yeh JL, et al. Subjective and objective sleep quality and aging in the sleep heart health study. *J Am Geriatr Soc* 2008 Jul;56(7):1218-1227. [doi: [10.1111/j.1532-5415.2008.01755.x](https://doi.org/10.1111/j.1532-5415.2008.01755.x)] [Medline: [18482295](https://pubmed.ncbi.nlm.nih.gov/18482295/)]
5. Minkel JD, Banks S, Htaik O, Moreta MC, Jones CW, McGlinchey EL, et al. Sleep deprivation and stressors: evidence for elevated negative affect in response to mild stressors when sleep deprived. *Emotion* 2012 Oct;12(5):1015-1020 [FREE Full text] [doi: [10.1037/a0026871](https://doi.org/10.1037/a0026871)] [Medline: [22309720](https://pubmed.ncbi.nlm.nih.gov/22309720/)]
6. Balter LJT, Sundelin T, Axelsson J. Sickness and sleep health predict frustration and affective responses to a frustrating trigger. *Sci Rep* 2021 Jan;11(1):1542 [FREE Full text] [doi: [10.1038/s41598-020-80461-4](https://doi.org/10.1038/s41598-020-80461-4)] [Medline: [33452313](https://pubmed.ncbi.nlm.nih.gov/33452313/)]
7. Saksvik-Lehouillier I, Saksvik SB, Dahlberg J, Tanum TK, Ringen HR, Karlsen H, et al. Mild to moderate partial sleep deprivation is associated with increased impulsivity and decreased positive affect in young adults. *Sleep* 2020 Oct;43(10):1-10 [FREE Full text] [doi: [10.1093/sleep/zsaa078](https://doi.org/10.1093/sleep/zsaa078)] [Medline: [32306048](https://pubmed.ncbi.nlm.nih.gov/32306048/)]
8. Steiger A, Kimura M. Wake and sleep EEG provide biomarkers in depression. *J Psychiatr Res* 2010 Mar;44(4):242-252. [doi: [10.1016/j.jpsychires.2009.08.013](https://doi.org/10.1016/j.jpsychires.2009.08.013)] [Medline: [19762038](https://pubmed.ncbi.nlm.nih.gov/19762038/)]
9. Pandi-Perumal SR, Monti JM, Burman D, Karthikeyan R, BaHammam AS, Spence DW, et al. Clarifying the role of sleep in depression: a narrative review. *Psychiatry Res* 2020 Sep;291:113239. [doi: [10.1016/j.psychres.2020.113239](https://doi.org/10.1016/j.psychres.2020.113239)] [Medline: [32593854](https://pubmed.ncbi.nlm.nih.gov/32593854/)]
10. Yang J, Fu X, Liao X, Li Y. Association of problematic smartphone use with poor sleep quality, depression, and anxiety: a systematic review and meta-analysis. *Psychiatry Res* 2020 Feb;284:112686. [doi: [10.1016/j.psychres.2019.112686](https://doi.org/10.1016/j.psychres.2019.112686)] [Medline: [31757638](https://pubmed.ncbi.nlm.nih.gov/31757638/)]
11. Chen Z, Lin M, Chen F, Lane ND, Cardone G, Wang R, et al. Unobtrusive sleep monitoring using smartphones. In: *Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare.*: IEEE; 2013 May Presented at: 7th International Conference on Pervasive Computing Technologies for Healthcare; May 5-8, 2013; Venice, Italy p. 145-152. [doi: [10.4108/icst.pervasivehealth.2013.252148](https://doi.org/10.4108/icst.pervasivehealth.2013.252148)]

12. Min JK, Doryab A, Wiese J, Amini S, Zimmerman J, Hong JI. Toss 'n' turn: smartphone as sleep and sleep quality detector. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York, NY, United States: Association for Computing Machinery; 2014 Apr Presented at: The 32nd Annual ACM Conference on Human Factors in Computing Systems; April 26-May 1, 2014; Toronto, Canada p. 477-486. [doi: [10.1145/2556288.2557220](https://doi.org/10.1145/2556288.2557220)]
13. Hoque E, Dickerson RF, Stankovic JA. Monitoring body positions and movements during sleep using WISPs. In: WH '10: Wireless Health 2010. New York, NY, United States: Association for Computing Machinery; 2010 Oct Presented at: Wireless Health 2010; October 5-7, 2010; San Diego, California p. 44-53. [doi: [10.1145/1921081.1921088](https://doi.org/10.1145/1921081.1921088)]
14. Montanini L, Sabino N, Spinsante S, Gambi E. Smartphone as unobtrusive sensor for real-time sleep recognition. : IEEE; 2018 Jan Presented at: 2018 IEEE International Conference on Consumer Electronics (ICCE); January 12-14, 2018; Las Vegas, United States p. 1-4. [doi: [10.1109/ICCE.2018.8326220](https://doi.org/10.1109/ICCE.2018.8326220)]
15. Sleep as Android. URL: <https://sleep.urbandroid.org/> [accessed 2022-02-24]
16. Sleep Cycle. URL: <https://www.sleepcycle.com/> [accessed 2022-02-24]
17. Lin YH, Wong BY, Lin SH, Chiu YC, Pan YC, Lee YH. Development of a mobile application (app) to delineate "digital chronotype" and the effects of delayed chronotype by bedtime smartphone use. J Psychiatr Res 2019 Mar;110:9-15. [doi: [10.1016/j.jpsychires.2018.12.012](https://doi.org/10.1016/j.jpsychires.2018.12.012)] [Medline: [30611008](https://pubmed.ncbi.nlm.nih.gov/30611008/)]
18. Abdullah S, Matthews M, Frank E, Doherty G, Gay G, Choudhury T. Automatic detection of social rhythms in bipolar disorder. J Am Med Inform Assoc 2016 May;23(3):538-543. [doi: [10.1093/jamia/ocv200](https://doi.org/10.1093/jamia/ocv200)] [Medline: [26977102](https://pubmed.ncbi.nlm.nih.gov/26977102/)]
19. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. New York, United States: Association for Computing Machinery; 2014 Sep Presented at: The 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle, United States p. 3-14. [doi: [10.1145/2632048.2632054](https://doi.org/10.1145/2632048.2632054)]
20. Huang K, Ding X, Xu J, Chen G, Ding W. Monitoring sleep and detecting irregular nights through unconstrained smartphone sensing. : IEEE; 2015 Aug Presented at: 2015 IEEE 12th Intl. Conf. on Ubiquitous Intelligence and Computing, 2015 IEEE 12th Intl. Conf. on Autonomic and Trusted Computing and 2015 IEEE 15th Intl. Conf. on Scalable Computing and Communications and its Associated Workshops (UIC-ATC-ScalCom); August 10-14, 2015; Beijing, China p. 36-45. [doi: [10.1109/UIC-ATC-ScalCom-CBDCCom-IoP.2015.30](https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCCom-IoP.2015.30)]
21. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: [10.1016/0165-1781\(89\)90047-4](https://doi.org/10.1016/0165-1781(89)90047-4)] [Medline: [2748771](https://pubmed.ncbi.nlm.nih.gov/2748771/)]
22. Bai Y, Xu B, Ma Y, Sun G, Zhao Y. Will you have a good sleep tonight? Sleep quality prediction with mobile phone. In: Proceedings of the 7th International Conference on Body Area Networks. Brussels, Belgium: ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering); 2012 Feb Presented at: 7th International Conference on Body Area Networks; September 24-26, 2012; Oslo, Norway p. 124-130. [doi: [10.4108/icst.bodynets.2012.250091](https://doi.org/10.4108/icst.bodynets.2012.250091)]
23. Natale V, Drejak M, Erbacci A, Tonetti L, Fabbri M, Martoni M. Monitoring sleep with a smartphone accelerometer. Sleep Biol Rhythms 2012 Oct;10(4):287-292. [doi: [10.1111/j.1479-8425.2012.00575.x](https://doi.org/10.1111/j.1479-8425.2012.00575.x)]
24. Bhat S, Ferraris A, Gupta D, Mozafarian M, DeBari VA, Gushway-Henry N, et al. Is there a clinical role for smartphone sleep apps? Comparison of sleep cycle detection by a smartphone application to polysomnography. J Clin Sleep Med 2015 Jul;11(7):709-715 [FREE Full text] [doi: [10.5664/jcsm.4840](https://doi.org/10.5664/jcsm.4840)] [Medline: [25766719](https://pubmed.ncbi.nlm.nih.gov/25766719/)]
25. Cuttone A, Bækgaard P, Sekara V, Jonsson H, Larsen JE, Lehmann S. Sensiblesleep: A Bayesian model for learning sleep patterns from smartphone events. PLoS One 2017 Jan;12(1):e0169901 [FREE Full text] [doi: [10.1371/journal.pone.0169901](https://doi.org/10.1371/journal.pone.0169901)] [Medline: [28076375](https://pubmed.ncbi.nlm.nih.gov/28076375/)]
26. Saeb S, Cybulski TR, Schueller SM, Kording KP, Mohr DC. Scalable passive sleep monitoring using mobile phones: opportunities and obstacles. J Med Internet Res 2017 Apr;19(4):e118 [FREE Full text] [doi: [10.2196/jmir.6821](https://doi.org/10.2196/jmir.6821)] [Medline: [28420605](https://pubmed.ncbi.nlm.nih.gov/28420605/)]
27. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. Increasing transparency through a multiverse analysis. Perspect Psychol Sci 2016 Sep;11(5):702-712. [doi: [10.1177/1745691616658637](https://doi.org/10.1177/1745691616658637)] [Medline: [27694465](https://pubmed.ncbi.nlm.nih.gov/27694465/)]
28. Van Baelen E, Geurts L, Van Waterschoot T, Mestdagh M, Kuppens P. Automatic Detection of Emotion and Emotion Disorder Using Smartphone Sensing. Leuven, Belgium: KU Leuven; 2018. URL: https://repository.teneo.libis.be/delivery/DeliveryManagerServlet?dps_pid=IE11053585&
29. de Beurs E, Van Dyck E, Marquenie LA, Lange A, Blonk RWB. De DASS: een vragenlijst voor het meten van depressie, angst en stress. Gedragstherapie 2001;34:35-53 [FREE Full text] [doi: [10.1037/t69646-000](https://doi.org/10.1037/t69646-000)]
30. 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](https://doi.org/10.1016/0005-7967(94)00075-U)] [Medline: [7726811](https://pubmed.ncbi.nlm.nih.gov/7726811/)]
31. Activity Recognition API. URL: <https://developers.google.com/location-context/activity-recognition> [accessed 2022-02-24]
32. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015 Jun;67(1):1-48 [FREE Full text] [doi: [10.18637/jss.v067.i01](https://doi.org/10.18637/jss.v067.i01)]

33. Schliep K, Hechenbichler K, Lizée A. kkn: Weighted k-nearest neighbors for classification, regression and clustering.: The Comprehensive R Archive Network; 2016 Mar 26. URL: <https://CRAN.R-project.org/package=kkn>
34. Karatzoglou A, Smola A, Hornik K, Zeileis A. kernlab - an S4 package for kernel methods in R. J Stat Soft 2004;11(9):1-20 [FREE Full text] [doi: [10.18637/jss.v011.i09](https://doi.org/10.18637/jss.v011.i09)]
35. Chen T, He T, Benesty M, Khotilovich V, Tang Y. xgboost: Extreme Gradient Boosting.: The Comprehensive R Archive Network; 2015. URL: <https://CRAN.R-project.org/package=xgboost> [accessed 2020-06-11]
36. Hastie T. gam: Generalized Additive Models.: The Comprehensive R Archive Network; 2020. URL: <https://CRAN.R-project.org/package=gam> [accessed 2020-06-11]
37. Kuhn M. Building predictive models in R using the caret package. J Stat Soft 2008 Nov;28(5):1-26 [FREE Full text] [doi: [10.18637/jss.v028.i05](https://doi.org/10.18637/jss.v028.i05)]
38. Liu F, Ting KM, Zhou Z. Isolation forest. : IEEE; 2008 Dec Presented at: 2008 Eighth IEEE International Conference on Data Mining; Pisa, Italy; December 15-19, 2008 p. 413-422. [doi: [10.1109/icdm.2008.17](https://doi.org/10.1109/icdm.2008.17)]
39. Cortes D. Distance approximation using isolation forests. arXiv 2019 Oct 27;arXiv:1910.12362 [stat.ML]:1-13 (forthcoming) [FREE Full text]
40. James G, Witten D, Hastie T, Tibshirani R. Cross-validation. In: Casella G, Fienberg S, Olkin I, editors. An Introduction to Statistical Learning. New York, United States: Springer-Verlag; 2013:176-178.
41. Breiman L, Friedman JH, Olshen RA, Stone CJ. Right sized trees and honest estimates. In: Classification And Regression Trees. New York, United States: Routledge; 1984:59-94.
42. Cawley GC, Nicola L, Talbot C. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res 2010 Oct;11:2079-2107 [FREE Full text]
43. Dejonckheere E, Mestdagh M, Houben M, Rutten I, Sels L, Kuppens P, et al. Complex affect dynamics add limited information to the prediction of psychological well-being. Nat Hum Behav 2019 May;3(5):478-491. [doi: [10.1038/s41562-019-0555-0](https://doi.org/10.1038/s41562-019-0555-0)] [Medline: [30988484](https://pubmed.ncbi.nlm.nih.gov/30988484/)]
44. Holm S. A simple sequentially rejective multiple test procedure. Scan J Stat 1979;6(2):65-70 [FREE Full text]

Abbreviations

AAR: Android activity recognition
ESM: experience sampling method
GAM: generalized additive model
kNN: k-nearest neighbors
LM: linear model
MAD: median absolute deviation
MLM: multilevel model
NA: negative affect
PSG: polysomnography
SSQ: subjective sleep quality
SVM: support vector machine
XGB: extreme gradient boosting

Edited by G Eysenbach; submitted 10.11.20; peer-reviewed by A Doryab, T Saheb, SH Lin; comments to author 15.03.21; revised version received 28.07.21; accepted 27.12.21; published 18.03.22.

Please cite as:

Niemeijer K, Mestdagh M, Kuppens P
Tracking Subjective Sleep Quality and Mood With Mobile Sensing: Multiverse Study
J Med Internet Res 2022;24(3):e25643
URL: <https://www.jmir.org/2022/3/e25643>
doi:[10.2196/25643](https://doi.org/10.2196/25643)
PMID:[35302502](https://pubmed.ncbi.nlm.nih.gov/35302502/)

©Koen Niemeijer, Merijn Mestdagh, Peter Kuppens. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 18.03.2022. 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.

Original Paper

Demographic and Psychosocial Characteristics Associated With Use of a Prostate Cancer Survivorship Website: Implications From a Multisite Randomized Controlled Trial

Allison Marziliano¹, PhD; Michael A Diefenbach¹, PhD; Shawna V Hudson², PhD; Erin K Tagai³, PhD, MPH; Elizabeth A Handorf³, PhD; Alicja Bator³, MPH; Suzanne M Miller³, PhD

¹Center for Health Innovations and Outcomes Research, Feinstein Institutes for Medical Research, Northwell Health, Manhasset, NY, United States

²Rutgers Robert Wood Johnson Medical School, Rutgers, The State University of New Jersey, New Brunswick, NJ, United States

³Cancer Prevention and Control, Fox Chase Cancer Center/Temple University Health System, Philadelphia, PA, United States

Corresponding Author:

Suzanne M Miller, PhD

Cancer Prevention and Control

Fox Chase Cancer Center/Temple University Health System

333 Cottman Avenue

Philadelphia, PA, 19111

United States

Phone: 1 1 215 728 4069

Email: Suzanne.Miller@fccc.edu

Abstract

Background: Many prostate cancer (PC) survivors experience disease and treatment-related symptomatology in both the physical and psychosocial domains. Although the benefits and barriers to using web-based resources for cancer patients are well-documented, less research has focused on the personal characteristics important for efficient tailoring and targeting of information that are associated with usage.

Objective: We used the Cognitive-Social Health Information Processing (C-SHIP) framework to guide our exploration of personal characteristics associated with use of PROGRESS, an informational PC survivorship website that addresses physical, emotional, interpersonal, and practical concerns relevant for PC survivors.

Methods: PC survivors (N=217) were randomized to the intervention arm (PROGRESS) of a randomized controlled trial. Of those randomized to the intervention arm, 84 used PROGRESS, and 133 did not use PROGRESS. Multivariable analyses evaluated demographic and psychosocial characteristics (eg, style of coping, health literacy, self-efficacy, affective states of depression, anxiety, and fatigue) associated with website use.

Results: A larger proportion of non-Hispanic White (68/160, 42.5%), compared with non-Hispanic Black (9/40, 23%), participants used PROGRESS ($P<.001$). Further, PROGRESS users were older in age ($P<.001$), had a monitoring style of coping ($P=.01$), and were less depressed ($P=.004$), anxious ($P=.02$), and fatigued ($P<.001$) than nonusers. Education, income, health literacy, blunting style of coping, self-efficacy, and treatment type (radiation therapy or surgery) were not significantly related to use. On multivariable analyses, race (OR 0.28, $P<.001$), age (OR 1.05, $P<.001$), monitoring style of coping (OR 1.27, $P=.02$), and overall mood (OR 0.98, $P<.001$) remained significant.

Conclusions: A combination of monitoring and low levels of negative affect were associated with website use. Additionally, users were older, non-Hispanic White survivors. To ensure that important survivorship-relevant information reaches users, future efforts need to focus on enhancing patient engagement.

Trial Registration: ClinicalTrials.gov NCT02224482; <https://clinicaltrials.gov/ct2/show/NCT02224482>

(*J Med Internet Res* 2022;24(3):e27890) doi:[10.2196/27890](https://doi.org/10.2196/27890)

KEYWORDS

prostate cancer; cancer survivorship; web-based resource; monitoring style of coping; cancer; survivorship; eHealth; emotions; interpersonal

Introduction

Prostate cancer (PC) is the second most common cancer diagnosed in men in the United States, with about 1 in 9 men diagnosed during their lifetime. The American Cancer Society estimates that, in 2021, there will be 248,530 new cases of PC in the United States [1]. The 5-year relative survival rate for localized PC is near 100% [1]. Consequently, the population of PC survivors is growing, exhibiting specific disease and treatment-related symptomatology in both the physical (eg, urinary, bowel, or sexual dysfunction [2,3]) and psychosocial (eg, high rates of distress, anxiety, reduced quality of life, depression, adjustment difficulties, fear of cancer recurrence [4-7]) domains. Further, studies have shown there is a significant interpersonal impact of PC on patients' relationships with their spouses and loved ones, as they struggle with intimacy, sexual confidence, sense of masculinity, familial cancer risk, and communicating about their diagnosis, treatment, and symptoms with friends [8,9].

The rapid development of modern technology has facilitated the use of web-based resources for individuals dealing with illness and treatment-related side effects, as is the case for PC survivors. Web-based resources have been developed and evaluated for many groups including breast cancer survivors [10,11], patients with melanoma [12], and families with parental cancer, as well as to educate nurses on reproductive issues in cancer patients [13]. Existing web-based resources specifically for patients with or survivors of PC include education or decision aids, interventions to reduce distress after treatment, and physical activity interventions [13-21].

Benefits to using web-based resources include ease of access at the patient's own schedule in a private place; ability to access the intervention through multiple channels (ie, personal computer, tablet, smartphone); augmented content through interactive videos, graphics, and testimonials; tailored content for treatment approaches or specific time points in the recovery trajectory; access transcending geographical barriers; and tracking patient recovery in real time [22-28]. Yet, there are also unique barriers, including lack of internet access, participants' privacy concerns, and program costs associated with developing and maintaining web-based resources [29-32]. In addition, studies have reported challenges engaging patients for initial access and staying engaged in recommended programs. A systematic review of adherence to web-based interventions showed that, on average, only about 50% of participants engage in software-based interventions, with variations from 10% to 90% across studies [33].

To increase engagement and persistent use of health-related software programs, it is therefore important to identify psychosocial characteristics that go beyond the commonly known access and demographic variables (eg, younger age, higher education) [28,29]. Thus, the current study aimed to identify patient demographic and psychosocial characteristics associated with use of PROGRESS, a web-based intervention for PC survivors.

Methods

Conceptual Framework

Our study is guided by the Cognitive-Social Health Information Processing (C-SHIP) model, a theoretical framework that identifies 5 cognitive-affective constructs that are associated with engagement in health protective behaviors [34,35]. These constructs consist of (1) cancer-relevant interpretations, (2) beliefs and expectations about cancer treatment and disease outcomes, (3) cancer-relevant goals and values, (4) cancer-relevant affective states, and (5) self-regulatory competencies and skills for generating and maintaining goal-oriented health-related behaviors. For the purpose of our analyses, these theoretical C-SHIP constructs were operationalized with the following patient-level variables: monitoring styles of coping (ie, the disposition to stand for and attend to health-relevant cues that entail C-SHIP's cancer-relevant interpretations, beliefs, and expectations about health risks); health literacy (C-SHIP's skills for generating and maintaining goal-oriented health behaviors); self-efficacy (C-SHIP's self-regulatory competencies); depression, anxiety, and fatigue (C-SHIP's cancer-relevant affective states). We hypothesized that these constructs are significantly and positively related to patient engagement and usage of the PROGRESS program.

Patient Recruitment

For the parent randomized controlled trial (RCT), PC patients were recruited during routine posttreatment follow-up appointments at 4 mid-Atlantic cancer centers. Recruitment occurred over the course of 3 years (2013-2016). Patients were eligible if they were diagnosed with localized PC (T1-T3c N0M0), were within 1 year of treatment completion, had regular access to a computer or a tablet with internet either at home or at another public place, were aged 18 years or older, were able to give consent, and were able to communicate in English. Exclusion criteria were presence of another primary cancer or a cancer recurrence.

Eligible patients who agreed to participate in the study were enrolled after signing the consent form and completing the baseline survey. Using block randomization by site, participants were randomized to either the control group (print materials: NCI's *Facing Forward: Life after Cancer Treatment* and *What You Need to Know about Prostate Cancer*) or the intervention group (PROGRESS + print materials). This manuscript focuses only on the subgroup of participants that were randomized to the intervention group, PROGRESS. This study was approved by the Institutional Review Boards at Fox Chase Cancer Center (#11 - 825), Rutgers University (#0220110092), Northwell Health (#14 - 672), and Mt. Sinai (#11 - 01136).

Intervention Condition

PROGRESS Content

PROGRESS is a self-paced, web-based educational program to address PC survivors' information needs in 6 specific domains suggested by prior work of the investigators [15,16,21,36-37], literature on this population, and formative work to develop the intervention content. These domains are (1) treatment type and

expected prostate specific antigen (PSA) changes, which indicate PC status and progression; (2) physical side effects (eg, urinary and sexual dysfunction); (3) emotional concerns (eg, fear of cancer recurrence); (4) interpersonal concerns (eg, communications with providers and family); (5) practical concerns (eg, follow-up care, financial needs); and (6) healthy lifestyle (eg, nutrition, physical activities) [15,38]. Information is culturally targeted and tailored to different survivorship stages (eg, short-term vs long-term survivorship needs). PROGRESS was developed through a 2-phase, qualitative formative research study. Phase 1 included individual interviews with 5 and group interviews with 12 early-stage prostate cancer patients to determine intervention content and interface. Phase 1 employed iterative user and usability testing (n=12) to finalize the intervention. Participants expressed interest in action-oriented content on managing treatment side effects, handling body image and comorbidities related to overweight or obesity, coping with emotional and communication issues, tips to reduce disruption of daily living activities, and health skills training tools. An extensive readability evaluation was conducted in order to ensure that the PROGRESS intervention met plain language standards. To conduct this evaluation, members of the Office of Health Communications and Health Disparities at Fox Chase Cancer Center used the software program, Health Literacy Advisor, to calculate reading grade levels and offer replacement text for complex terms and long sentences. All text was revised as needed to conform to an 8th or 9th grade target level reading range. For more information on the development, preliminary testing, and efficacy of PROGRESS, see Miller et al [38] and Tagai et al [39,40].

PROGRESS features include a topics tab (addressing financial or legal issues; interpersonal communication, emotional and practical concerns; negative feelings; and side effects); videos from physicians, patients, and content experts; fields for personal tracking (of PSA level, health status, weight, sleep, urinary or erectile dysfunction, medication, living habits, and questions for upcoming physician appointments); information on the latest PC findings (in prevention, screening, treatment, and survivorship); a virtual health center and navigator; theoretically guided normalizing messages; testimonials from a group of diverse PC survivors; and technology support (a tutorial program and a help desk). The software program underwent extensive usability testing before it was released for study purposes [38].

Translating Theory-Driven Constructs to Practical Content

The 5 key theoretical constructs were operationalized within PROGRESS through the program's components of (1) providing accurate information, (2) creating realistic expectations and promoting self-efficacy, (3) exploring the patient's goals and values and encouraging behavior consistent with them, (4) validating feelings and facilitating emotional support, and (5) providing information and training to maximize self-regulatory competencies and skills.

Data Collection

Assessments were conducted at baseline and at 1-month, 3-month, and 6-month follow-ups. Data for these analyses were drawn from the baseline assessment and indication of website

use during the study. Patients completed the assessments via paper-and-pencil survey, online via REDCap, or telephone interview. Follow-up telephone calls or emails (based on participant's preferred survey completion method) were used for noncompleted surveys. Participants received a US \$20 gift card for each completed assessment.

Study Measures

Demographics assessed at baseline included race/ethnicity, age, and education. Comorbidities were captured using the Charlson Comorbidity Index, a 16-item weighted measure evaluating the presence or absence, or severity, of illnesses [41]. The Charlson is scored as a weighted sum of the illnesses, such that a positive response for certain illnesses (ie, myocardial infarction, congestive health failure, dementia) adds 1 point each, a positive response for other illnesses (ie, leukemia) adds 2 points each, and a positive response for AIDS adds 6 points. The remaining illnesses, which are scored on a severity scale, are multiplied by 2, 3, 2, and 6 for diabetes mellitus, liver disease, renal disease, and malignant solid tumor, respectively. Style of coping was assessed using the Monitor/Blunter Style Scale (MBSS) [42]. Health Literacy was assessed using a 3-item screen for health literacy [43,44]. The 3 items are: "How often do you have someone help you read hospital materials?", "How confident are you filling out medical forms by yourself?", and "How often do you have problems learning about your medical condition because of difficulty understanding written information?" Response options are on a 5-point Likert scale ranging from "always to never" or "extremely to not at all." Self-efficacy for symptom control was measured using an author-constructed 12-item scale that asked participants how confident they were to "manage any treatment-related fatigue," "recover your emotional well-being," and "does your family know how to support you?" [45]. Depression, anxiety, and fatigue were assessed with the Profile of Mood States Short Form (POMS-SF), using the respective depression, anxiety, and fatigue subscales [46]. Depression was also assessed with the Center for Epidemiological Studies-Depression subscale (CES-D) [47]. Undergoing surgery or radiation therapy was assessed via self-report and confirmed with medical chart abstraction.

The outcome variable, PROGRESS use, was a binary variable coded as "use" or "nonuse" and was obtained via Google Analytics after completion of the participants' last follow-up assessments. Participants randomized to PROGRESS were considered to have used PROGRESS if they clicked beyond the home page at least once during the study period. Participants were categorized as nonusers if they did not log in or logged in but did not click through to any other page beyond the home page. We were unable to track amount of website use with the available tracking metrics.

Statistical Analyses

Data were analyzed with SPSS version 19.0 and R version 3.6. Descriptive statistics were used to characterize the sample [48,49]. We then used a series of univariable logistic regression models to test the association between website use and each variable of interest, using generalized estimating equations with robust standard errors to account for within-site correlation. We

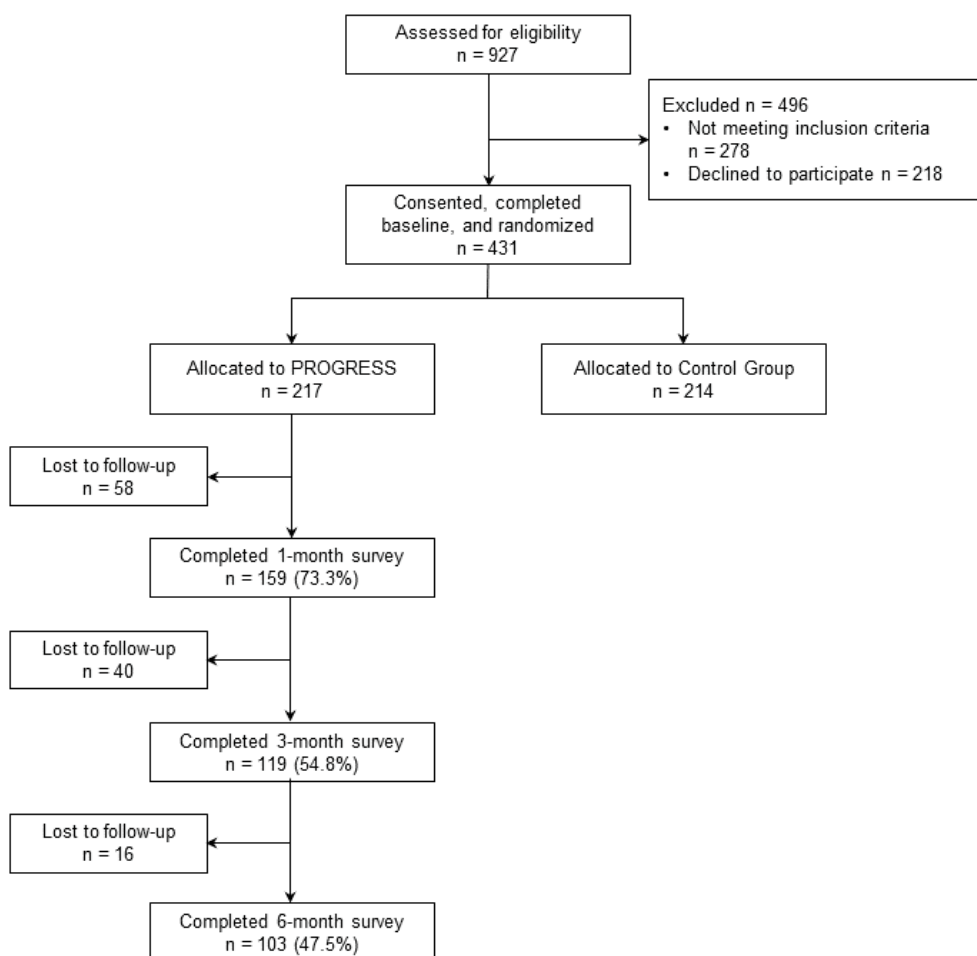
also used a multivariable logistic regression model to simultaneously evaluate the variables' associations with website use. Significant variables from the univariate logistic regression analyses, a priori variables of interest, and surgery or radiation treatment (hormone therapy was excluded due to low number of patients receiving this treatment) were included in the multivariable logistic regression model. To avoid collinearity, the POMS-SF total score was included rather than the individual subscales. A 2-sided *P* value <.05 was considered statistically significant.

Results

Recruitment

A total of 927 participants were assessed for eligibility; 278 did not meet inclusion criteria, leaving 649 eligible. A total of 431 participants (66.4% of those eligible) consented and were enrolled and randomized (217 PC survivors were randomized to PROGRESS, and 214 were randomized to the control condition), and 218 declined to participate. Of those randomized to PROGRESS, 73.3% (159/217) completed the 1-month time point, 54.8% (119/217) completed the 3-month time point, and 47.5% (103/217) completed the 6-month time point. See [Figure 1](#) for a detailed CONSORT (Consolidated Standards of Reporting Trials) diagram, adapted for this study focusing on the PROGRESS intervention arm.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) diagram for the PROGRESS intervention arm.



Sample Demographics

Overall, the sample was mostly non-Hispanic White (160/217, 73.7%), married (175/217, 80.6%), and with no comorbidities (167/214, 78.0%). About half (103/215, 47.9%) had a college degree or higher, and about half (115/217, 53.0%) endorsed that they were in the highest income bracket (US \$75,001 and greater). In terms of treatment, about half (112/217, 51.6%) had surgery. The average age of the sample was 63.79 (SD 6.67) years. See [Table 1](#) for details.

Participants' mean scores were 2.46 (SD 2.20) and 1.44 (SD 1.41) on the monitoring and blunting subscales, respectively; 13.12 (SD 2.24) on the measure of health literacy; 8.75 (SD 1.11) on the self-efficacy for re-entry scale; 4.76 (SD 4.32) on the anxiety measure; and 6.27 (SD 5.05) on the measure of fatigue. Further, the average scores on the POMS-SF depression measure and CES-D measure were 3.30 (SD 3.94) and 5.99 (SD 5.27), respectively. See [Table 1](#) for details.

Table 1. Demographic characteristics of website users and nonusers (controlling for study site).

Variable	Website users (n=84)	Nonusers (n=133)	Total sample (N=217)	P value
Race/ethnicity, n (%)				
Non-Hispanic White	68 (81.0)	92 (69.2)	160 (73.7)	ref ^a
Non-Hispanic Black	9 (10.7)	31 (23.3)	40 (18.4)	<.001
All other races	7 (8.3)	10 (7.5)	17 (7.8)	.86
Age (years), mean (SD)	65.37 (7.03)	62.79 (6.27)	63.79 (6.67)	<.001
Education^b, n (%)				
High school or less	16 (19.0)	33 (25.2)	49 (22.8)	ref
Some college	26 (31.0)	37 (28.2)	63 (29.3)	.40
College degree	22 (26.2)	33 (25.2)	55 (25.6)	.08
Graduate/professional degree	20 (23.8)	28 (21.4)	48 (22.3)	.60
Household income (US \$), n (%)				
<45,000	11 (13.1)	25 (18.8)	36 (16.6)	ref
45,001-75,000	21 (25.0)	31 (23.3)	52 (24.0)	.34
≥75,001	49 (58.3)	66 (49.6)	115 (53.0)	.25
Missing	3 (3.6)	11 (8.3)	14 (6.5)	.30
Marital status, n (%)				
Never married, divorced, separated, widowed, single, unknown, refused	12 (14.3)	30 (22.6)	42 (19.4)	ref
Married or domestic partnership	72 (85.7)	103 (77.4)	175 (80.6)	.08
Health literacy, mean (SD)	13.51 (1.94)	12.87 (2.39)	13.12 (2.24)	.18
Blunting, mean (SD)	1.65 (1.24)	1.31 (1.50)	1.44 (1.41)	.29
Monitoring, mean (SD)	3.01 (1.93)	2.11 (2.29)	2.46 (2.20)	.01
Self-efficacy for re-entry, mean (SD)	8.83 (0.98)	8.70 (1.18)	8.75 (1.11)	.07
Depression (POMS-SF ^c), mean (SD)	2.89 (3.25)	3.56 (4.31)	3.30 (3.94)	.15
Depression (CES-D ^d), mean (SD)	5.29 (4.17)	6.43 (5.84)	5.99 (5.27)	<.001
Tense/anxiety, mean (SD)	4.36 (3.74)	5.01 (4.65)	4.76 (4.32)	.02
Fatigue, mean (SD)	5.72 (4.64)	6.61 (5.28)	6.27 (5.05)	<.001
Charlson Comorbidity Index^e, n (%)				
0	66 (79.5)	101 (77.1)	167 (78.0)	ref
1	10 (12.1)	18 (13.7)	28 (13.1)	.52
≥2	7 (8.4)	12 (9.2)	19 (8.9)	.30
Radiation therapy, n (%)				
No	52 (61.9)	96 (72.2)	148 (68.2)	.09
Yes	32 (38.1)	37 (27.8)	69 (31.8)	
Surgery, n (%)				
No	39 (46.4)	66 (49.6)	105 (48.4)	.80
Yes	45 (53.6)	67 (50.4)	112 (51.6)	

^aref: reference.^bTotal n=215.^cPOMS-SF: Profile of Mood States Short Form.^dCES-D: Center for Epidemiological Studies-Depression subscale.^eTotal n=214.

Main Analyses

Of the 217 patients, 84 (38.7%) reported using the website versus the 133 (61.3%) who reported that they did not use the website. When controlling for study site, there were significant differences between those who used PROGRESS and those who did not use PROGRESS in the following variables: race/ethnicity, age, style of coping, depression, anxiety, and fatigue. Specifically, a larger proportion of non-Hispanic White (68/160, 42.5%), compared with non-Hispanic Black (9/40, 23%), participants used PROGRESS ($P<.001$; Table 1). Users of PROGRESS were older in age (mean 65.37, SD 7.03 years vs mean 62.79, SD 6.27 years; $P<.001$), higher on monitoring (mean 3.01, SD 1.93 vs mean 2.11, SD 2.29; $P=.01$), and less depressed (mean 5.29, SD 4.17 vs mean 6.43, SD 5.84; $P<.001$), anxious (mean 4.36, SD 3.74 vs mean 5.01, SD 4.65; $P=.02$),

and fatigued (mean 5.72, SD 4.64 vs mean 6.61, SD 5.28; $P<.001$). There were no other significant differences between those who used PROGRESS and those who did not use PROGRESS.

In the multivariable model, non-Hispanic Black participants were significantly less likely to use the website than non-Hispanic White participants (OR 0.28, 95% CI 0.25-0.32; Table 2). There was no significant difference between non-Hispanic White participants and those of all other race/ethnicities (OR 0.77, 95% CI 0.43-1.36). Additionally, older participants were more likely to use the website (OR 1.05, 95% CI 1.04-1.07), as well as those reporting greater monitoring style of coping (OR 1.27, 95% CI 1.04-1.56). Finally, those with a more negative mood state were significantly less likely to use the website (OR 0.98, 95% CI 0.97-0.98).

Table 2. Multivariable logistic regression analysis (controlling for study site).

Variable	OR ^a	SE	LCL ^b	UCL ^c
Non-Hispanic Black ^d	0.28	.07	0.25	0.32
All other races/ethnicities ^d	0.77	.29	0.43	1.36
Age	1.05	.01	1.04	1.07
Monitoring	1.27	.10	1.04	1.56
Self-efficacy for re-entry	0.89	.07	0.78	1.01
Mood total	0.98	.004	0.97	0.98
Radiation therapy ^e	1.42	.19	0.98	2.07
Surgery ^e	1.19	.39	0.56	2.57

^aOR: odds ratio.

^bLCL: lower confidence limit.

^cUCL: upper confidence limit.

^dReference group: non-Hispanic White.

^eReference group: did not receive treatment.

Discussion

Principal Findings

PROGRESS was more likely to be used by PC survivors who were high on monitoring style of coping, a coping style to deal with threat that involves scanning for and magnifying disease-related cues [42]. Monitors are often more concerned about their illness, experience more treatment-related side effects, are more knowledgeable about their medical situation, feel themselves to be at greater personal risk, and are less satisfied with and more demanding about the psychosocial aspects of their care [50]. Notably, they are more often adherent to medical recommendations and place greater value on health-related information [50]. The PROGRESS website offered patients characterized as monitors authoritative information about PC from providers and testimonials by a diverse group of survivors who are addressing psychosocial aspects of the disease. Thus, PROGRESS offers monitors health-related information that they typically value and that translated into higher use.

Results also showed that PROGRESS was more likely to be used by non-Hispanic, White PC survivors, confirming prior established patterns of internet use. A study conducted by the Pew Research Center indicated that internet usage is more common in White compared with non-Hispanic Black populations [51,52]. Racial disparities in education are well-documented and favor non-minority populations [53,54]. Patients with higher levels of education are more likely to have used a computer in the past and would be more apt to use offered resources, such as PROGRESS.

PROGRESS users' relatively positive mood (less depressed, less anxious, and less fatigued) allows them to mobilize in support of their health, which includes taking advantage of the information PROGRESS has to offer. Using PROGRESS and other related health resources may be blunted in patients who are highly depressed, as depressed patients often cannot empower themselves to take care of their health. Similarly, high anxiety may hinder the use of PROGRESS, as these patients may have anxiety about the information that PROGRESS will provide. Patients with high levels of fatigue may not feel energized to a degree necessary to use a web-based resource such as PROGRESS. Our results also showed that older patients

were more likely to use PROGRESS. Though this finding was statistically significant, it was not clinically significant, with an average age of 65.37 years compared with 62.79 years. It is encouraging that older men are using this web-based resource, as PC survivors more generally are an older demographic given the average age of diagnosis is 66 years old [1]. Analyzing age as a categorical variable produces the same results. It is possible that the older participants were more motivated to use the site or that this finding is due to chance, as the numbers for each individual group are not that large.

The study findings highlight the need for strategies to increase patient engagement with web-based tools, as less than 40% of the intervention group reported that they used PROGRESS. Engagement needs to go beyond the commonly accepted development and access strategies, such as ensuring access to a computer and the internet (eg, providing tablets in clinic during down time), ensuring comfort with using a computer and the internet, using language and terminology appropriate for a low health-literate population, incorporating culturally targeted material into the program, and prompting use through text message or email reminders. The value of the recommended services needs to be clearly communicated, and if possible, gamification elements, such as badges or virtual competition with other users, can be incorporated.

Indeed, our study team did engage in a thoughtful, iterative process to design the PROGRESS website that included stakeholder feedback. Based on the initial design phase and the usability testing, the PROGRESS site was very positively evaluated. We think that this highlights the need for ongoing usability and acceptability testing, rather than collecting these data at one time point prior to intervention launch. This experience has suggested that, for future studies, researchers should build in a regular review of usage, usability, and acceptability and devise a plan for how to handle the responses if certain intervention components are not well-received. Collecting these data longitudinally during the RCT phase may not be able to inform the intervention being tested, but it would be useful to inform future interventions.

Low patient engagement is a threat to efforts to evaluate the efficacy of web-based tools. Before beginning a research study, power calculations are needed to ascertain sample size requirements necessary to detect clinically meaningful differences. Understanding that all of those enrolled in the intervention arm may not actually engage with the intervention may alter researchers' plans for how many patients to enroll. On a related note, researchers may, a priori, plan to conduct their analyses in 2 ways: first, comparing intervention and control, and second, comparing within the intervention arm, users with nonusers. In addition, rather than "using website" as a simple binary variable, future research should employ more sophisticated website use tracking features that allow investigators to capture detailed website use, such as time spent with each page, pages with most views, and number of downloads.

Limitations

This study is not without its limitations. First, the sample was highly homogenous and was mostly non-Hispanic White and high income, and thus, the results may not be generalizable to other, more diverse patient groups. Second, our "website use" variable was binary and therefore did not allow us to evaluate the full continuum of website usage from those who never logged in to the super-users who used PROGRESS over and over again. These limitations aside, we believe our study substantially contributes to the literature characterizing the patient profile of a web-based resource user.

Conclusions

Our study showed that, compared with nonusers, users of PROGRESS, a website for PC survivors, were more likely to be non-Hispanic White (compared with non-Hispanic Black participants), be older in age, have a higher monitoring style of coping, and have experienced higher levels of positive mood. Improved engagement features need to be developed and evaluated to increase the perceived value to patients. Additionally, the existence of a user patient profile indicates the potential to tailor web-based resources accordingly.

Acknowledgments

The study was funded by the National Cancer Institute (R01CA224918, P30CA006927, P30CA072720, and T32CA009035), the National Center for Advancing Translational Sciences (UL1TR003017), the Centers for Disease Control and Prevention (U01OH011690), and the American Cancer Society (RSG-15-021-01-CPPB).

Conflicts of Interest

None declared.

References

1. Key Statistics in Prostate Cancer. American Cancer Society. URL: <https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html> [accessed 2020-10-18]
2. Ávila M, Patel L, López S, Cortés-Sanabria L, Garin O, Pont À, et al. Patient-reported outcomes after treatment for clinically localized prostate cancer: A systematic review and meta-analysis. *Cancer Treat Rev* 2018 May;66:23-44. [doi: [10.1016/j.ctrv.2018.03.005](https://doi.org/10.1016/j.ctrv.2018.03.005)] [Medline: [29673922](https://pubmed.ncbi.nlm.nih.gov/29673922/)]

3. Chen C, Chen Z, Wang K, Hu L, Xu R, He X. Comparisons of health-related quality of life among surgery and radiotherapy for localized prostate cancer: a systematic review and meta-analysis. *Oncotarget* 2017 Nov 17;8(58):99057-99065 [FREE Full text] [doi: [10.18632/oncotarget.21519](https://doi.org/10.18632/oncotarget.21519)] [Medline: [29228751](https://pubmed.ncbi.nlm.nih.gov/29228751/)]
4. De Sousa A, Sonavane S, Mehta J. Psychological aspects of prostate cancer: a clinical review. *Prostate Cancer Prostatic Dis* 2012 Jun;15(2):120-127. [doi: [10.1038/pcan.2011.66](https://doi.org/10.1038/pcan.2011.66)] [Medline: [22212706](https://pubmed.ncbi.nlm.nih.gov/22212706/)]
5. Sharpley CF, Bitsika V, Christie DRH. "The Worst Thing Was...": prostate cancer patients' evaluations of their diagnosis and treatment experiences. *Am J Mens Health* 2018 Sep;12(5):1503-1509 [FREE Full text] [doi: [10.1177/1557988318772752](https://doi.org/10.1177/1557988318772752)] [Medline: [29708020](https://pubmed.ncbi.nlm.nih.gov/29708020/)]
6. Deimling GT, Bowman KF, Sterns S, Wagner LJ, Kahana B. Cancer-related health worries and psychological distress among older adult, long-term cancer survivors. *Psychooncology* 2006 Apr;15(4):306-320. [doi: [10.1002/pon.955](https://doi.org/10.1002/pon.955)] [Medline: [16041841](https://pubmed.ncbi.nlm.nih.gov/16041841/)]
7. Bellizzi KM, Latini DM, Cowan JE, DuChane J, Carroll PR. Fear of recurrence, symptom burden, and health-related quality of life in men with prostate cancer. *Urology* 2008 Dec;72(6):1269-1273. [doi: [10.1016/j.urology.2007.12.084](https://doi.org/10.1016/j.urology.2007.12.084)] [Medline: [18342930](https://pubmed.ncbi.nlm.nih.gov/18342930/)]
8. Campbell LC, Keefe FJ, Scipio C, McKee DC, Edwards CL, Herman SH, et al. Facilitating research participation and improving quality of life for African American prostate cancer survivors and their intimate partners. A pilot study of telephone-based coping skills training. *Cancer* 2007 Jan 15;109(2 Suppl):414-424 [FREE Full text] [doi: [10.1002/cncr.22355](https://doi.org/10.1002/cncr.22355)] [Medline: [17173280](https://pubmed.ncbi.nlm.nih.gov/17173280/)]
9. Harden J, Northouse L, Cimprich B, Pohl JM, Liang J, Kershaw T. The influence of developmental life stage on quality of life in survivors of prostate cancer and their partners. *J Cancer Surviv* 2008 Jun;2(2):84-94 [FREE Full text] [doi: [10.1007/s11764-008-0048-z](https://doi.org/10.1007/s11764-008-0048-z)] [Medline: [18648977](https://pubmed.ncbi.nlm.nih.gov/18648977/)]
10. Irene Su H, Stark S, Kwan B, Boles S, Chingos D, Ehren J, et al. Efficacy of a web-based women's health survivorship care plan for young breast cancer survivors: a randomized controlled trial. *Breast Cancer Res Treat* 2019 Aug;176(3):579-589 [FREE Full text] [doi: [10.1007/s10549-019-05260-6](https://doi.org/10.1007/s10549-019-05260-6)] [Medline: [31054032](https://pubmed.ncbi.nlm.nih.gov/31054032/)]
11. Ridner SH, Dietrich MS, Davis AJ, Sinclair V. A randomized clinical trial comparing the impact of a web-based multimedia intervention versus an educational pamphlet on patient outcomes in breast cancer survivors with chronic secondary lymphedema. *J Womens Health (Larchmt)* 2020 May;29(5):734-744 [FREE Full text] [doi: [10.1089/jwh.2019.7676](https://doi.org/10.1089/jwh.2019.7676)] [Medline: [31314661](https://pubmed.ncbi.nlm.nih.gov/31314661/)]
12. Coups EJ, Manne SL, Ohman Strickland P, Hilgart M, Goydos JS, Heckman CJ, et al. Randomized controlled trial of the mySmartSkin web-based intervention to promote skin self-examination and sun protection behaviors among individuals diagnosed with melanoma: study design and baseline characteristics. *Contemp Clin Trials* 2019 Aug;83:117-127 [FREE Full text] [doi: [10.1016/j.cct.2019.06.014](https://doi.org/10.1016/j.cct.2019.06.014)] [Medline: [31255801](https://pubmed.ncbi.nlm.nih.gov/31255801/)]
13. Quinn GP, Bowman Curci M, Reich RR, Gwede CK, Meade CD, ENRICH/ECHO Working Group, et al. Impact of a web-based reproductive health training program: ENRICH (Educating Nurses about Reproductive Issues in Cancer Healthcare). *Psychooncology* 2019 May;28(5):1096-1101 [FREE Full text] [doi: [10.1002/pon.5063](https://doi.org/10.1002/pon.5063)] [Medline: [30882960](https://pubmed.ncbi.nlm.nih.gov/30882960/)]
14. Baptista S, Teles Sampaio E, Heleno B, Azevedo LF, Martins C. Web-based versus usual care and other formats of decision aids to support prostate cancer screening decisions: systematic review and meta-analysis. *J Med Internet Res* 2018 Jun 26;20(6):e228 [FREE Full text] [doi: [10.2196/jmir.9070](https://doi.org/10.2196/jmir.9070)] [Medline: [29945857](https://pubmed.ncbi.nlm.nih.gov/29945857/)]
15. Diefenbach MA, Benedict C, Miller SM, Stanton AL, Ropka ME, Wen K, et al. Examining the impact of a multimedia intervention on treatment decision-making among newly diagnosed prostate cancer patients: results from a nationwide RCT. *Transl Behav Med* 2018 Nov 21;8(6):876-886 [FREE Full text] [doi: [10.1093/tbm/iby066](https://doi.org/10.1093/tbm/iby066)] [Medline: [29982747](https://pubmed.ncbi.nlm.nih.gov/29982747/)]
16. Diefenbach MA, Mohamed NE, Butz BP, Bar-Chama N, Stock R, Cesaretti J, et al. Acceptability and preliminary feasibility of an internet/CD-ROM-based education and decision program for early-stage prostate cancer patients: randomized pilot study. *J Med Internet Res* 2012 Jan 13;14(1):e6 [FREE Full text] [doi: [10.2196/jmir.1891](https://doi.org/10.2196/jmir.1891)] [Medline: [22246148](https://pubmed.ncbi.nlm.nih.gov/22246148/)]
17. Ellison GL, Weinrich SP, Lou M, Xu H, Powell IJ, Baquet CR. A randomized trial comparing web-based decision aids on prostate cancer knowledge for African-American men. *J Natl Med Assoc* 2008 Oct;100(10):1139-1145 [FREE Full text] [doi: [10.1016/s0027-9684\(15\)31481-4](https://doi.org/10.1016/s0027-9684(15)31481-4)] [Medline: [18942274](https://pubmed.ncbi.nlm.nih.gov/18942274/)]
18. Tomko C, Davis KM, Luta G, Krist AH, Woolf SH, Taylor KL. A comparison of web-based versus print-based decision AIDS for prostate cancer screening: participants' evaluation and utilization. *J Gen Intern Med* 2015 Jan;30(1):33-42 [FREE Full text] [doi: [10.1007/s11606-014-2994-7](https://doi.org/10.1007/s11606-014-2994-7)] [Medline: [25183475](https://pubmed.ncbi.nlm.nih.gov/25183475/)]
19. Cockle-Hearne J, Barnett D, Hicks J, Simpson M, White I, Faithfull S. A web-based intervention to reduce distress after prostate cancer treatment: development and feasibility of the getting down to coping program in two different clinical settings. *JMIR Cancer* 2018 Apr 30;4(1):e8 [FREE Full text] [doi: [10.2196/cancer.8918](https://doi.org/10.2196/cancer.8918)] [Medline: [29712628](https://pubmed.ncbi.nlm.nih.gov/29712628/)]
20. Golsteijn RHJ, Bolman C, Peels DA, Volders E, de Vries H, Lechner L. A web-based and print-based computer-tailored physical activity intervention for prostate and colorectal cancer survivors: a comparison of user characteristics and intervention use. *J Med Internet Res* 2017 Aug 23;19(8):e298 [FREE Full text] [doi: [10.2196/jmir.7838](https://doi.org/10.2196/jmir.7838)] [Medline: [28835353](https://pubmed.ncbi.nlm.nih.gov/28835353/)]
21. Diefenbach MA, Butz BP. A multimedia interactive education system for prostate cancer patients: development and preliminary evaluation. *J Med Internet Res* 2004 Jan 21;6(1):e3 [FREE Full text] [doi: [10.2196/jmir.6.1.e3](https://doi.org/10.2196/jmir.6.1.e3)] [Medline: [15111269](https://pubmed.ncbi.nlm.nih.gov/15111269/)]

22. Chung C, Cooper SJ, Cant RP, Connell C, McKay A, Kinsman L, et al. The educational impact of web-based and face-to-face patient deterioration simulation programs: An interventional trial. *Nurse Educ Today* 2018 May;64:93-98. [doi: [10.1016/j.nedt.2018.01.037](https://doi.org/10.1016/j.nedt.2018.01.037)] [Medline: [29459198](https://pubmed.ncbi.nlm.nih.gov/29459198/)]
23. Liaw SY, Chng DYJ, Wong LF, Ho JTY, Mordiffi SZ, Cooper S, et al. The impact of a Web-based educational program on the recognition and management of deteriorating patients. *J Clin Nurs* 2017 Dec;26(23-24):4848-4856. [doi: [10.1111/jocn.13955](https://doi.org/10.1111/jocn.13955)] [Medline: [28685966](https://pubmed.ncbi.nlm.nih.gov/28685966/)]
24. Sowan AK, Idhail JA. Evaluation of an interactive web-based nursing course with streaming videos for medication administration skills. *Int J Med Inform* 2014 Aug;83(8):592-600. [doi: [10.1016/j.ijmedinf.2014.05.004](https://doi.org/10.1016/j.ijmedinf.2014.05.004)] [Medline: [24930589](https://pubmed.ncbi.nlm.nih.gov/24930589/)]
25. Hallett J, Maycock B, Kypri K, Howat P, McManus A. Development of a Web-based alcohol intervention for university students: processes and challenges. *Drug Alcohol Rev* 2009 Jan;28(1):31-39 [FREE Full text] [doi: [10.1111/j.1465-3362.2008.00008.x](https://doi.org/10.1111/j.1465-3362.2008.00008.x)] [Medline: [19320673](https://pubmed.ncbi.nlm.nih.gov/19320673/)]
26. McRee A, Shoben A, Bauermeister JA, Katz ML, Paskett ED, Reiter PL. Outsmart HPV: Acceptability and short-term effects of a web-based HPV vaccination intervention for young adult gay and bisexual men. *Vaccine* 2018 Dec 18;36(52):8158-8164 [FREE Full text] [doi: [10.1016/j.vaccine.2018.01.009](https://doi.org/10.1016/j.vaccine.2018.01.009)] [Medline: [29331245](https://pubmed.ncbi.nlm.nih.gov/29331245/)]
27. Pugatch J, Grenen E, Surla S, Schwarz M, Cole-Lewis H. Information architecture of web-based interventions to improve health outcomes: systematic review. *J Med Internet Res* 2018 Mar 21;20(3):e97 [FREE Full text] [doi: [10.2196/jmir.7867](https://doi.org/10.2196/jmir.7867)] [Medline: [29563076](https://pubmed.ncbi.nlm.nih.gov/29563076/)]
28. Stawarz K, Preist C, Coyle D. Use of smartphone apps, social media, and web-based resources to support mental health and well-being: online survey. *JMIR Ment Health* 2019 Jul 12;6(7):e12546 [FREE Full text] [doi: [10.2196/12546](https://doi.org/10.2196/12546)] [Medline: [31301126](https://pubmed.ncbi.nlm.nih.gov/31301126/)]
29. Speed E, Davison C, Gunnell C. The anonymity paradox in patient engagement: reputation, risk and web-based public feedback. *Med Humanit* 2016 Jun;42(2):135-140. [doi: [10.1136/medhum-2015-010823](https://doi.org/10.1136/medhum-2015-010823)] [Medline: [26879526](https://pubmed.ncbi.nlm.nih.gov/26879526/)]
30. Hilton JF, Barkoff L, Chang O, Halperin L, Ratanawongsa N, Sarkar U, et al. A cross-sectional study of barriers to personal health record use among patients attending a safety-net clinic. *PLoS One* 2012 Feb 20;7(2):e31888 [FREE Full text] [doi: [10.1371/journal.pone.0031888](https://doi.org/10.1371/journal.pone.0031888)] [Medline: [22363761](https://pubmed.ncbi.nlm.nih.gov/22363761/)]
31. Brown LL, Lustria MLA, Rankins J. A review of web-assisted interventions for diabetes management: maximizing the potential for improving health outcomes. *J Diabetes Sci Technol* 2007 Nov;1(6):892-902 [FREE Full text] [doi: [10.1177/193229680700100615](https://doi.org/10.1177/193229680700100615)] [Medline: [19885163](https://pubmed.ncbi.nlm.nih.gov/19885163/)]
32. Heiniger LE, Smith AB, Olver I, Grimison P, Klein B, Wootten A, et al. e-TC: Development and pilot testing of a web-based intervention to reduce anxiety and depression in survivors of testicular cancer. *Eur J Cancer Care (Engl)* 2017 Nov;26(6):1. [doi: [10.1111/ecc.12698](https://doi.org/10.1111/ecc.12698)] [Medline: [28544085](https://pubmed.ncbi.nlm.nih.gov/28544085/)]
33. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEW. Persuasive system design does matter: a systematic review of adherence to web-based interventions. *J Med Internet Res* 2012 Nov 14;14(6):e152 [FREE Full text] [doi: [10.2196/jmir.2104](https://doi.org/10.2196/jmir.2104)] [Medline: [23151820](https://pubmed.ncbi.nlm.nih.gov/23151820/)]
34. Miller SM, Shoda Y, Hurley K. Applying cognitive-social theory to health-protective behavior: breast self-examination in cancer screening. *Psychol Bull* 1996 Jan;119(1):70-94. [doi: [10.1037/0033-2909.119.1.70](https://doi.org/10.1037/0033-2909.119.1.70)] [Medline: [8559860](https://pubmed.ncbi.nlm.nih.gov/8559860/)]
35. Miller SM, Tagai EK, Wen K, Lee M, Hui SA, Kurtz D, et al. Predictors of adherence to follow-up recommendations after an abnormal Pap smear among underserved inner-city women. *Patient Educ Couns* 2017 Jul;100(7):1353-1359 [FREE Full text] [doi: [10.1016/j.pec.2017.01.020](https://doi.org/10.1016/j.pec.2017.01.020)] [Medline: [28190541](https://pubmed.ncbi.nlm.nih.gov/28190541/)]
36. Marcus AC, Diefenbach MA, Stanton AL, Miller SM, Fleisher L, Raich PC, CISRC Research Team. Cancer patient and survivor research from the cancer information service research consortium: a preview of three large randomized trials and initial lessons learned. *J Health Commun* 2013;18(5):543-562 [FREE Full text] [doi: [10.1080/10810730.2012.743629](https://doi.org/10.1080/10810730.2012.743629)] [Medline: [23448232](https://pubmed.ncbi.nlm.nih.gov/23448232/)]
37. Wen K, Miller SM, Stanton AL, Fleisher L, Morra ME, Jorge A, et al. The development and preliminary testing of a multimedia patient-provider survivorship communication module for breast cancer survivors. *Patient Educ Couns* 2012 Aug;88(2):344-349 [FREE Full text] [doi: [10.1016/j.pec.2012.02.003](https://doi.org/10.1016/j.pec.2012.02.003)] [Medline: [22770812](https://pubmed.ncbi.nlm.nih.gov/22770812/)]
38. Miller SM, Hudson SV, Hui SA, Diefenbach MA, Fleisher L, Raivitch S, et al. Development and preliminary testing of PROGRESS: a Web-based education program for prostate cancer survivors transitioning from active treatment. *J Cancer Surviv* 2015 Sep;9(3):541-553 [FREE Full text] [doi: [10.1007/s11764-015-0431-5](https://doi.org/10.1007/s11764-015-0431-5)] [Medline: [25697335](https://pubmed.ncbi.nlm.nih.gov/25697335/)]
39. Tagai EK, Hudson SV, Diefenbach MA, Xu J, Bator A, Marziliano A, et al. Social and medical risk factors associated with supportive needs in the first year following localized prostate cancer treatment. *J Cancer Surviv* 2021 Feb 18;15(1):110-118 [FREE Full text] [doi: [10.1007/s11764-020-00916-5](https://doi.org/10.1007/s11764-020-00916-5)] [Medline: [32681305](https://pubmed.ncbi.nlm.nih.gov/32681305/)]
40. Tagai EK, Miller SM, Hudson SV, Diefenbach MA, Handorf E, Bator A, et al. Improved cancer coping from a web-based intervention for prostate cancer survivors: A randomized controlled trial. *Psychooncology* 2021 Sep;30(9):1466-1475. [doi: [10.1002/pon.5701](https://doi.org/10.1002/pon.5701)] [Medline: [33855796](https://pubmed.ncbi.nlm.nih.gov/33855796/)]
41. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis* 1987;40(5):373-383. [doi: [10.1016/0021-9681\(87\)90171-8](https://doi.org/10.1016/0021-9681(87)90171-8)] [Medline: [3558716](https://pubmed.ncbi.nlm.nih.gov/3558716/)]

42. Miller SM. Monitoring and blunting: validation of a questionnaire to assess styles of information seeking under threat. *J Pers Soc Psychol* 1987 Feb;52(2):345-353. [doi: [10.1037//0022-3514.52.2.345](https://doi.org/10.1037//0022-3514.52.2.345)] [Medline: [3559895](https://pubmed.ncbi.nlm.nih.gov/3559895/)]
43. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. *J Gen Intern Med* 1995 Oct;10(10):537-541. [doi: [10.1007/BF02640361](https://doi.org/10.1007/BF02640361)] [Medline: [8576769](https://pubmed.ncbi.nlm.nih.gov/8576769/)]
44. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med* 2004 Sep;36(8):588-594 [FREE Full text] [Medline: [15343421](https://pubmed.ncbi.nlm.nih.gov/15343421/)]
45. Anderson KO, Dowds BN, Pelletz RE, Edwards WT, Peeters-Asdourian C. Development and initial validation of a scale to measure self-efficacy beliefs in patients with chronic pain. *Pain* 1995 Oct;63(1):77-83. [doi: [10.1016/0304-3959\(95\)00021-J](https://doi.org/10.1016/0304-3959(95)00021-J)] [Medline: [8577493](https://pubmed.ncbi.nlm.nih.gov/8577493/)]
46. Curran SL, Andrykowski MA, Studts JL. Short Form of the Profile of Mood States (POMS-SF): Psychometric information. *Psychological Assessment* 1995;7(1):80-83. [doi: [10.1037/1040-3590.7.1.80](https://doi.org/10.1037/1040-3590.7.1.80)]
47. Radloff LS. The CES-D Scale. *Applied Psychological Measurement* 2016 Jul 26;1(3):385-401. [doi: [10.1177/014662167700100306](https://doi.org/10.1177/014662167700100306)]
48. IBM SPSS software. IBM Corp. URL: <https://www.ibm.com/analytics/spss-statistics-software> [accessed 2022-02-12]
49. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing. 2013. URL: <http://www.R-project.org/> [accessed 2022-02-12]
50. Roussi P, Miller SM. Monitoring style of coping with cancer related threats: a review of the literature. *J Behav Med* 2014 Oct;37(5):931-954 [FREE Full text] [doi: [10.1007/s10865-014-9553-x](https://doi.org/10.1007/s10865-014-9553-x)] [Medline: [24488543](https://pubmed.ncbi.nlm.nih.gov/24488543/)]
51. Internet/Broadband Fact Sheet. Pew Research Center. 2019. URL: <https://www.pewresearch.org/internet/fact-sheet/internet-broadband/> [accessed 2020-10-10]
52. Internet Health Resources. Pew Research Center. 2003 Jul 16. URL: <https://www.pewresearch.org/internet/2003/07/16/internet-health-resources/> [accessed 2020-12-10]
53. Bacharach VR, Baumeister AA, Furr RM. Racial and gender science achievement gaps in secondary education. *J Genet Psychol* 2003 Mar;164(1):115-126. [doi: [10.1080/00221320309597507](https://doi.org/10.1080/00221320309597507)] [Medline: [12693748](https://pubmed.ncbi.nlm.nih.gov/12693748/)]
54. Voight A, Hanson T, O'Malley M, Adekanye L. The racial school climate gap: within-school disparities in students' experiences of safety, support, and connectedness. *Am J Community Psychol* 2015 Dec;56(3-4):252-267. [doi: [10.1007/s10464-015-9751-x](https://doi.org/10.1007/s10464-015-9751-x)] [Medline: [26377419](https://pubmed.ncbi.nlm.nih.gov/26377419/)]

Abbreviations

CES-D: Center for Epidemiological Studies-Depression subscale

CONSORT: Consolidated Standards of Reporting Trials

C-SHIP: Cognitive-Social Health Information Processing

MBSS: Monitor/Blunter Style Scale

PC: prostate cancer

POMS-SF: Profile of Mood States Short Form

PROGRESS: Prostate Cancer Survivorship Website

PSA: prostate specific antigen

RCT: randomized controlled trial

Edited by R Kukafka; submitted 12.02.21; peer-reviewed by C Pimentel, J Rowland; comments to author 19.05.21; revised version received 13.09.21; accepted 14.10.21; published 21.03.22.

Please cite as:

Marziliano A, Diefenbach MA, Hudson SV, Tagai EK, Handorf EA, Bator A, Miller SM

Demographic and Psychosocial Characteristics Associated With Use of a Prostate Cancer Survivorship Website: Implications From a Multisite Randomized Controlled Trial

J Med Internet Res 2022;24(3):e27890

URL: <https://www.jmir.org/2022/3/e27890>

doi: [10.2196/27890](https://doi.org/10.2196/27890)

PMID: [35311678](https://pubmed.ncbi.nlm.nih.gov/35311678/)

©Allison Marziliano, Michael A Diefenbach, Shawna V Hudson, Erin K Tagai, Elizabeth A Handorf, Alicja Bator, Suzanne M Miller. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 21.03.2022. 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.

Original Paper

Validity Evidence of the eHealth Literacy Questionnaire (eHLQ) Part 2: Mixed Methods Approach to Evaluate Test Content, Response Process, and Internal Structure in the Australian Community Health Setting

Christina Cheng^{1,2*}, PhD; Gerald R Elsworth^{1,2*}, PhD; Richard H Osborne^{1,2*}, PhD

¹School of Health Sciences, Centre for Global Health and Equity, Swinburne University of Technology, Hawthorn, Australia

²Faculty of Health, School of Health and Social Development, Deakin University, Burwood, Australia

*all authors contributed equally

Corresponding Author:

Christina Cheng, PhD

School of Health Sciences

Centre for Global Health and Equity

Swinburne University of Technology

Room 907, Level 9, AMCD Building

453/469-477 Burwood Road

Hawthorn, 3122

Australia

Phone: 61 392145470

Email: cccheng@swin.edu.au

Abstract

Background: Digital technologies have changed how we manage our health, and eHealth literacy is needed to engage with health technologies. Any eHealth strategy would be ineffective if users' eHealth literacy needs are not addressed. A robust measure of eHealth literacy is essential for understanding these needs. On the basis of the eHealth Literacy Framework, which identified 7 dimensions of eHealth literacy, the eHealth Literacy Questionnaire (eHLQ) was developed. The tool has demonstrated robust psychometric properties in the Danish setting, but validity testing should be an ongoing and accumulative process.

Objective: This study aims to evaluate validity evidence based on test content, response process, and internal structure of the eHLQ in the Australian community health setting.

Methods: A mixed methods approach was used with cognitive interviewing conducted to examine evidence on test content and response process, whereas a cross-sectional survey was undertaken for evidence on internal structure. Data were collected at 3 diverse community health sites in Victoria, Australia. Psychometric testing included both the classical test theory and item response theory approaches. Methods included Bayesian structural equation modeling for confirmatory factor analysis, internal consistency and test-retest for reliability, and the Bayesian multiple-indicators, multiple-causes model for testing of differential item functioning.

Results: Cognitive interviewing identified only 1 confusing term, which was clarified. All items were easy to read and understood as intended. A total of 525 questionnaires were included for psychometric analysis. All scales were homogenous with composite scale reliability ranging from 0.73 to 0.90. The intraclass correlation coefficient for test-retest reliability for the 7 scales ranged from 0.72 to 0.95. A 7-factor Bayesian structural equation modeling using small variance priors for cross-loadings and residual covariances was fitted to the data, and the model of interest produced a satisfactory fit (posterior productive $P=.49$, 95% CI for the difference between observed and replicated chi-square values -101.40 to 108.83 , prior-posterior productive $P=.92$). All items loaded on the relevant factor, with loadings ranging from 0.36 to 0.94. No significant cross-loading was found. There was no evidence of differential item functioning for administration format, site area, and health setting. However, discriminant validity was not well established for scales 1, 3, 5, 6, and 7. Item response theory analysis found that all items provided precise information at different trait levels, except for 1 item. All items demonstrated different sensitivity to different trait levels and represented a range of difficulty levels.

Conclusions: The evidence suggests that the eHLQ is a tool with robust psychometric properties and further investigation of discriminant validity is recommended. It is ready to be used to identify eHealth literacy strengths and challenges and assist the development of digital health interventions to ensure that people with limited digital access and skills are not left behind.

(*J Med Internet Res* 2022;24(3):e32777) doi:[10.2196/32777](https://doi.org/10.2196/32777)

KEYWORDS

eHealth; health literacy; health equity; questionnaire design; validity evidence; eHLQ; mobile phone

Introduction

Background

Digital technologies have brought fundamental changes to modern-day life including how we manage our health. We can quickly search for health information at our fingertips but are also facing an avalanche of misinformation, as evident during the COVID-19 pandemic [1]. We can have instant access to our electronic personal health record, but these digital systems can be difficult to use or do not meet our expectation [2-5]. In addition, some people simply do not have or only have limited access or skills to use technologies for health, leading to the potential widening of health inequity when people with limited access or skills are being left behind in the digital age.

To characterize the challenges of accessing and using digital technologies for health, the concept of eHealth literacy was coined in 2006 [6]. At that time, it was defined as “the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem” [6]. It is further recognized that any digital or eHealth strategy or intervention would be ineffective if the eHealth literacy needs of users are not addressed [6-8]. For example, in the postevaluation phase of the now defunct web-based personal health record in the United Kingdom, the HealthSpace, it was concluded that there was a mismatch of the system and users’ expectations, and some users seemed to lack the health literacy and digital literacy required to use the system [2,3].

The eHealth Literacy Questionnaire

To describe and address eHealth literacy needs, Norgaard et al [9] developed the eHealth Literacy Framework using a grounded validity-driven approach based on the results of a series of concept mapping workshops with a diverse range of stakeholders. Therefore, 7 dimensions of eHealth literacy were identified, depicting an overarching vision of how people access, understand, and use technology for health involving skills, health systems, and interaction between the individual and the systems [9]. This grounded framework provides a more comprehensive and contemporary view of eHealth literacy than the original definition, as it also taps into the role of digital systems and the interaction between users and systems. The framework was subsequently used to develop the eHealth Literacy Questionnaire (eHLQ) as a tool to measure eHealth literacy based on the seven dimensions:

1. Using technology to process health information
2. Understanding of health concepts and language
3. Ability to actively engage with digital services
4. Feel safe and in control

5. Motivated to engage with digital services
6. Access to digital services that work
7. Digital services that suit individual needs [10]

With the inclusion of eHealth literacy dimensions relating to user interaction and experiences in using digital health systems, the eHLQ embraces the real-world experiences of users while capturing the interactivity and increasing capabilities of digital technologies. It can provide rich information about the competencies of individuals as well as the maturity of digital health systems, as mature systems are likely to be more responsive to the individual needs of users [10]. As such, the eHLQ is a useful tool for digital health developers and implementers to promote equity-driven digital health systems and eventually contribute to health equity.

The eHLQ was simultaneously developed in Danish and English to avoid idiom and improve item wording to enhance future translation of the questionnaire into other languages [10]. The tool consists of 35 items with 7 scales, each with 4 to 6 items, and the response option is a 4-point ordinal scale of *strongly disagree* to *strongly agree*. The results are 7 equally weighted composite scale scores [10]. Initial validity testing of the Danish version of the eHLQ, based on data collected from 475 people from community and health care settings in Denmark, demonstrated the psychometric robustness of the tool. A 7-factor confirmatory factor analysis (CFA) model using the Bayesian structural equation modeling (BSEM) approach resulted in a satisfactory fit. All the 35 items of the tool loaded strongly on their relevant factors (range 0.36-0.94) with no significant cross-loadings. Good internal consistency was also demonstrated with satisfactory composite scale reliability for each of the 7 scales (range 0.75-0.87). In addition to taking the classical test theory (CTT) approach to the testing of psychometric properties, the item response theory (IRT) approach was also used. The results confirmed that the 35 items represented a range of difficulties and had good discrimination for testing people with different levels of eHealth literacy ability. Measurement invariance for age and sex was also demonstrated [10]. Although the study provided satisfactory validity evidence of the eHLQ in the Danish setting, evidence for the English version needs to be established. Described as a pioneer of eHealth in the world, Denmark has a national, publicly owned eHealth portal that is used by at least 2.3 million unique users out of their 5.8 million citizens per month as of 2019 [11,12]. Digital health is part of the routine for many Danish citizens. Although Australia also has a national digital health record system, information from an Australian Senate estimates hearing in 2019 revealed that only 4% of Australians had logged in to the Australian system [13]. Therefore, how the eHLQ will perform in settings with less prominent public use of digital health services is not known.

Validity Evidence

Validity testing is an ongoing process that involves the accumulation of 5 sources of evidence based on test content, response process, internal structure, relations to other variables, and consequences of testing, according to the authoritative reference of developing and using of educational and psychological measurements, the *Standards for Educational and Psychological Testing* (the *Standards*) [14].

Evidence based on test content is used to determine whether the items represent the content domain and may also include whether the wordings are easy to read and formats of administration are easy to use. Response process refers to the cognitive process of survey participants, that is, whether the interpretation of the items by participants aligns with the intended interpretation of items by test developers. It may also include whether interpretation remains the same across subgroups or across different formats of administration. Internal structure is the extent to which items conform to the constructs and relates to aspects such as factor analysis, reliability, and measurement invariance. Relations to other variables is the analysis of the relationship between the scores on another instrument relevant in the theoretical network of the construct being measured or other external variables that the scores can predict, whereas consequences of testing relates to the robustness of the proposed use of the test scores, including intended benefits, indirect effects, and unintended consequences such as construct underrepresentation or construct irrelevance [14-19]. This study focuses on the evidence collected in the Australian community health setting. Evidence on relations to other variables in this setting has been described by Cheng et al [20]. The aim of this study is to report and evaluate the evidence on test content, response process, and internal structure of the English eHLQ in Australia.

Methods

Data Collection

Methods to collect and evaluate the validity evidence were guided by the discussion in the *Standards* and related literature. A mixed methods approach was used with cognitive interviewing conducted to examine evidence on test content and response process, whereas a cross-sectional survey was undertaken for evidence on internal structure. Cognitive interviews were conducted in 2017 at a not-for-profit community health organization in the metropolitan area of Victoria, Australia. The clients of this site, together with clients from 2 private primary care medical clinics in the metropolitan and regional areas, were invited to participate in the cross-sectional survey in 2018.

Eligibility criteria for participation in both activities were clients aged ≥ 18 years, with or without any health conditions, and able to complete the questionnaire in paper-based format, web-based format, or face-to-face interview. Clients experiencing significant cognitive or mental health issues, who were too clinically unwell, and with insufficient English to complete the questionnaire and who did not have a carer to assist them were excluded.

Ethics Approval

Ethical approval of the study was obtained from the Deakin University Human Research Ethics Committee (HEAG-H 146_2017).

Cognitive Interviewing

Cognitive interviewing is commonly used to explore the cognitive process of how people answer survey items [21-23]. It may shed light on how people construct their answers to determine if their thinking matches the item as intended by the test developers, if people experience difficulties when answering the questions, or if the layouts are suitable. The results can also be used to identify response differences across sociocultural groups [17,22,24,25].

Given the qualitative nature of cognitive interviewing, a large sample size is not required but needs to be representative and diverse [14,24]. The process took an iterative approach that involved rounds of testing should issues be identified and the questions needed to be revised [21], with all items tested at least five times or until data saturation [26,27].

Participants were recruited with assistance from the health site, and a plain language information sheet was provided, with written consent requested. Interviews were conducted after participants completed the paper-based format. Participant behavior was observed when they answered the questionnaire. Upon completion, two questions were asked to gain insights into the cognitive process: (1) What were you thinking when you answered this question? (2) Why did you choose this answer? Participants were encouraged to make any further comments about the items or the format. They could be interviewed for all 35 items or part of the questionnaire, mainly for older participants to avoid fatigue and cognitive overload.

Data analysis was conducted using text summary [24]. Content analysis was first reviewed against the item intents of the eHLQ by one of the authors (CC), and further review was undertaken by another author (RHO). Any issues identified for revisions were discussed among all the authors until agreement was reached.

Cross-sectional Survey for Psychometric Testing

For the cross-sectional survey, clients were recruited at the waiting area and provided with an information sheet. A signed consent form was not requested with the return of the completed questionnaire as implied consent as a strategy to facilitate participation. Apart from self-administration using paper-based or web-based formats, interviews were included to ensure that older people or people with low literacy could participate. Demographic questions including age, sex, postcode, language spoken at home, education, health status, perceived health status, and use of digital services were also collected. Further description of recruitment is described in the study by Cheng et al [20].

Similar to the Danish eHLQ validity testing [10], this study also adopted both the CTT and IRT approaches for psychometric analysis. CTT is the traditional approach based on the assertion that an observed score comprises a true score and an error score [28,29]. This approach usually involves the evaluation of

dimensionality, discrimination, and reliability. However, CTT has been criticized for being sample dependent and does not take into account the characteristics of test items and how people at different levels of the construct of interest perform on those items, which is the focus of IRT, the modern approach [28,29]. Hence, both approaches were used in this study to strengthen the collection of evidence.

Statistical Analysis

Overview

Analyses were conducted using three statistical software programs, namely, SPSS (version 25.0; IBM Corp) [30], Mplus (version 8.3; Muthén & Muthén) [31], and IRTPRO (Item Response Theory for Patient-Reported Outcomes; version 4.20; Vector Psychometric Group) [32]. Descriptive statistics were generated for the demographic data, eHLQ scores, and floor and ceiling effects. The presence of floor and ceiling effects may indicate poor discrimination at the minimum or maximum values [33], with effects considered significant if over 15% of participants score in the top (ceiling) or bottom (floor) of a score range [34,35].

Missing Values

To deal with missing data for the eHLQ scores, the data set was first examined. If no clear pattern of missingness was found, that is, the missingness could be regarded as completely at random, a 2-step approach would be taken. The first step was to delete cases with more than 50% of missing values to reduce potential bias. The second step was to replace all missing values using the expectation-maximization algorithm imputation in SPSS [30,36,37]. This final data set was used for all psychometric analyses.

CTT Analysis

Item difficulty is an item property and is usually conducted as a first step in item analysis in the CTT approach [28,38]. This parameter was calculated as the proportion of survey participants who endorsed disagree and strongly disagree against the proportion of agree and strongly agree [39]. Hence, the higher the proportion responding to disagree and strongly disagree indicated a higher level of difficulty.

To measure reliability, internal consistency and test-retest reliability were evaluated. In addition to the commonly used Cronbach α for internal consistency, which has been criticized for producing biased estimates when items do not have equal factor loadings or in case of correlated item errors [39], composite scale reliability calculated through structural equation modeling using Mplus recommended by Raykov [40] was also evaluated. These 2 estimates are expected to be fairly comparable when the set of items is unidimensional, has uncorrelated errors, and has high loadings on the true score [29,40]. The acceptable range of both estimates is from 0.70 to 0.95, with 0.80 generally regarded as good reliability [41]. For test-retest, there is no consensus on the optimal length between time points, and invitations were sent 1 week after the first completion of the questionnaire. Test-retest reliability was evaluated using the intraclass correlation coefficient (ICC) [42]. A minimum of 10 participants is considered adequate for

detecting an acceptable ICC of 0.70 with 80% power at a significance level of .05 [43], and a sample size of 30 was estimated.

Following the classical item analysis, CFA was conducted, given that the hypothesized constructs were specified a priori. Similar to the Danish validity testing, the BSEM approach was adopted [10] using Mplus. There is no agreement on the sample size for CFA, which can range from 100 to 400 [33,44], and a sample size of 500 was estimated. Contrary to the traditional frequentist approach using the maximum likelihood estimation procedure, which assumes that measures not related to a latent factor will have zero loadings on that factor, which can easily lead to poor model fit and rejection [29,45], the parameters in BSEM are treated as variables, and this more flexible approach is described as a better reflection of substantive theories [45].

The different parameter specifications in BSEM at the start of an analysis are described as priors, which can be diffuse (noninformative) or informative [45]. Diffuse priors are hypothesized parameters that are fully estimated from the data, whereas informative priors are likely parameter values derived from previous studies, researchers' theories, or prior beliefs [45]. Informative priors can be applied to cross-loadings and residual covariances. For cross-loading priors, a variance prior of 0.01 means that 95% of the loading variation is within the range of -0.20 to $+0.20$, which is considered a small loading. Inverse-Wishart priors were used for the covariances among item residuals. The application is to start with a large enough df of the inverse-Wishart distribution and gradually lowers the df parameter to produce a more flexible model such that the residual covariances are not strictly constrained to zero [46].

A sequence of 1-factor models followed by a 7-factor model (Multimedia Appendix 1) were fitted to the data. Several models with different parameter specifications, using the Danish study as a reference [10], were tested and compared to identify the model of interest, which is a model that is not rejected by the data and is "closest to the CFA model that fits well enough" [46]. To evaluate model fit, Mplus produces a posterior predictive P (PPP) value and 95% CI for the difference between the observed and replicated chi-square values. A low PPP value and positive 95% CI indicate poor fit, whereas a PPP value of around .5 and a value of 0 falling close to the middle of the 95% CI indicates excellent fit [45]. Furthermore, a nonsignificant prior-posterior predictive P (PPPP) value, that is, $>.05$, indicates that the estimates of the cross-loadings can be considered approximate zero and are thus negligible [47]. Apart from the aforementioned estimates, a model comparison can be conducted by examining the model convergence and information criterion. The model of interest is the model with quicker convergence, that is, when the potential scale reduction is consistently less than 1.05 by the smallest number of iterations, and has a lower discrepancy information criterion [46].

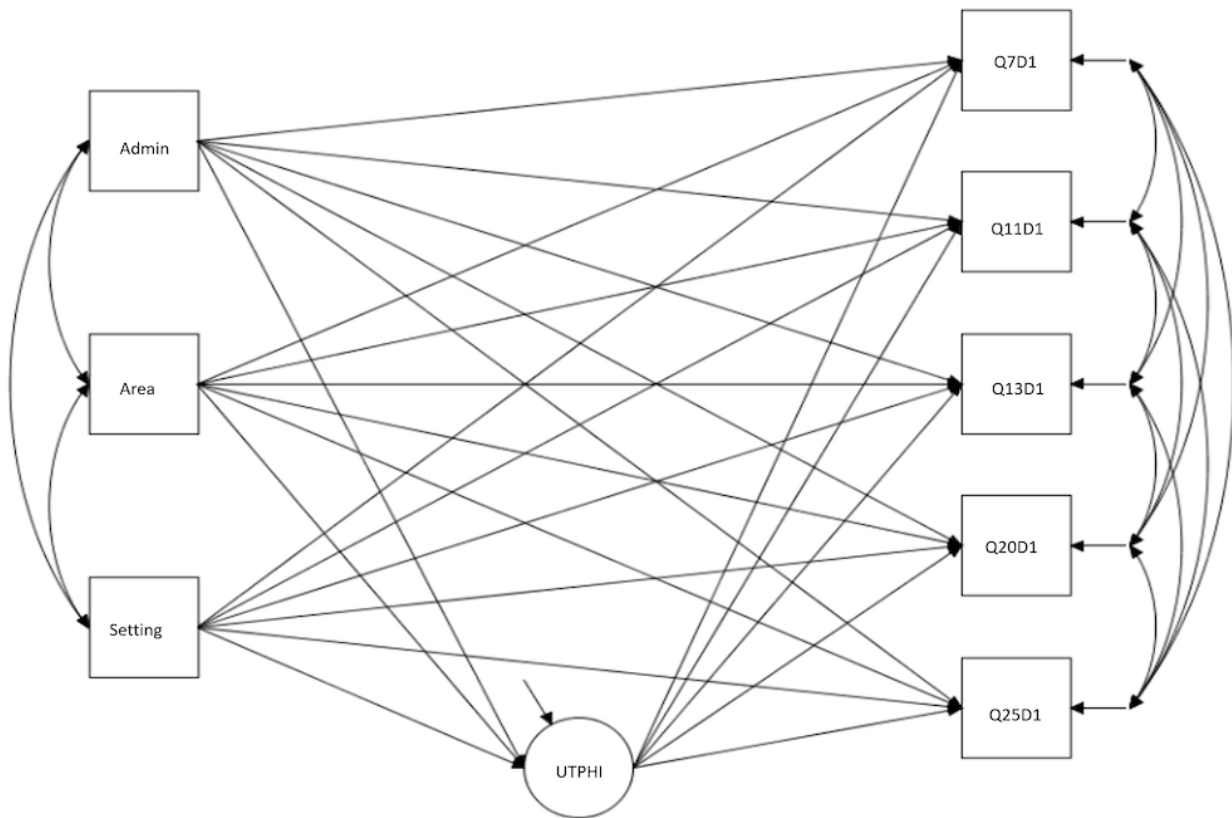
On the basis of the results of the selected 7-factor model, discriminant validity was evaluated using the Fornell-Larcker criteria [48] based on shared variance and average variance extracted (AVE). Shared variance is calculated by squaring the interfactor correlation between 2 factors, and the AVE is generated by the average of the sum of the squared factor

loadings of the related items [49]. If the AVE is >0.5, the first Fornell-Larcker criterion that “a factor accounts for more variance in the observed variables associated with it than measurement error or similar external, unmeasured influences” [49] is met. To meet the second criterion about the factor’s association with other factors in the conceptual framework, the AVE of any 2 factors both “have to be greater than the shared variance estimate” between the 2 factors [48,49].

The BSEM approach is further used to test for differential item functioning (DIF), that is, the stability of measurement across different groups or grouping variables [50,51]. By using the multiple-indicators, multiple-causes (MIMIC) model approach, both DIF and group differences can be detected simultaneously [52]. This method has been used to test for group differences among different demographic groups in the same setting and is described with details in the study by Cheng et al [20]. For this analysis, administration format (paper-based vs face-to-face

interviews), site area (metropolitan vs regional area), and health setting (private clinic vs not-for-profit community health) were included as covariates. The web-based format was not included for analysis owing to the limited number of participants (13/525, 2.5%) using this format. See Figure 1 for the Bayesian model for testing DIF with scale 1 as an example. Procedures for model testing and selection are similar to BSEM described earlier, except priors were applied to DIF paths instead of cross-loadings. For the evaluation of DIF, a significant direct effect indicated the presence of DIF, and a 1-tailed *P* value of <.025 was considered significant as no directional hypotheses were set up. If DIF was identified, the estimates were further evaluated based on the model selected. For example, if the model with variance prior of 0.01 was selected and the *PPPP* value was nonsignificant, then estimates within the range of -0.2 to +0.2 could be considered negligible [53]. Group differences were also examined as supplementary results.

Figure 1. Bayesian multiple-indicators, multiple-causes model for differential item functioning testing with scale 1 of eHealth Literacy Questionnaire as the example. Output from Mplus: Admin: administration format (0=face-to-face interview, 1=paper format); Area: site area, that is, location of participating organization (0=metropolitan, 1=regional); Setting: health setting (0=private clinic, 1=community health); UTPHI: eHealth Literacy Questionnaire scale 1 (using technology to process health information); Q7D1, Q11D1, Q13D1, Q20D1, and Q25D1: eHealth Literacy Questionnaire items.



IRT Analysis

To perform an IRT analysis, 4 assumptions need to be met. The assumptions of unidimensionality (items are measuring the same construct), local independence (each item should not be related except they are measuring the same construct), and item invariance (item parameters are the same across subgroups) can be examined through the CTT methods described in the CTT Analysis section. The assumption of monotonicity (the probability of endorsing an item increases as the trait level

increases) is evaluated by examining the test characteristic curves [54]. The sample size requirement for IRT analysis may range from 200 to 500 [54]. For this analysis, the generalized partial credit model, similar to the Danish study, was applied using IRTPRO. Apart from the test characteristic curves, item thresholds, item location, item discrimination, and information functions from the 7 unidimensional IRT models were evaluated. Item response thresholds were evaluated by inspecting the item characteristic curves, where the peak of each response category curve from the lowest (strongly disagree) to the highest (strongly

agree) should correspond to the lowest to the highest trait level to demonstrate that an item has an ordered set of response thresholds. For item discrimination, a steeper curve or slope of the item characteristic curves indicates better discrimination between people with different levels of the trait. A higher item discrimination estimate indicates higher discrimination between people with differences in ability [55,56]. For item location estimates, lower estimates represent easier items, as these items are expected to be endorsed by people with lower ability and are expected to be similar to the item difficulty estimates in the CTT analysis [54,57,58]. Finally, an inspection of the information function curve provided information on where an item could precisely measure the underlying trait. This measurement precision is analogous to the reliability in CTT [14]. It also helps to determine if the items of a scale measure the full spectrum of the underlying trait.

Results

Cognitive Interviewing

A total of 12 participants were recruited for 2 rounds of cognitive interviews. Of these 12 participants, 8 (67%) were women and 4 (33%) were men, with 58% (7/12) of the participants aged >65 years and 17% (2/12) speaking a language other than English at home. The sample provided a good representation of people from a range of different age groups and cultural backgrounds. The first round with 7 participants identified the term *health technology services* used in the 4 items of scale 7 (*Digital services that suit individual needs*) was confusing. Most participants could not immediately relate the term to digital systems as intended. Following the discussion within the research team, the term was changed to *eHealth systems*. The changed term was tested in a second round with

5 participants, and no further issue was identified. Participants' understanding of the other items was generally similar to the item intents. For example, for items about using *technology* to find, understand, share, or organize health information, participants could easily link it to the internet, Dr Google, or anything web based. In the item "I use measurements about my body....," participants thought about how they used results of blood test, body weight, or blood pressure, which was aligned with the item intent.

Despite the diverse backgrounds of participants, no major differences in understanding the items were identified, and all participants found the items easy to read. Recommendations from participants also led to changes in the introductory page to provide examples of technology, health technology, eHealth systems, and health care providers or health professionals. The completion time of the questionnaire ranged from 3 to <7 minutes.

Psychometric Testing

Participant Characteristics

A total of 530 completed questionnaires were collected. On the basis of the treatment of missing values described in the Statistical Analysis section, 5 cases were deleted, leading to a final sample size of 525 for psychometric analyses. The age of participants of the cross-sectional survey ranged from 18 to 94 years, and 61% (320/525) of the participants were women. A total of 33.3% (175/525) of the participants had a university education, and 30.9% (162/525) spoke a language other than English at home. Of the 525 participants, 300 (57.1%) reported having some form of chronic illness. Regarding technology use, of the 525 participants, 151 (28.8%) did not have a computer or laptop, and 131 (25%) did not use email or SMS text messaging (Table 1).

Table 1. Characteristics of cross-sectional survey participants (N=525).

Characteristics	Value
Age (years), mean (SD; range)	56.8 (18.6; 18-94)
Setting, n (%)	
Site 1: metropolitan private clinic	204 (38.9)
Site 2: metropolitan community health	204 (38.9)
Site 3: regional private clinic	117 (22.3)
Administration format, n (%)	
Paper-based	399 (76)
Web-based	13 (2.5)
Face-to-face interview	113 (21.5)
Sex, n (%)	
Female	320 (61)
Male	203 (38.7)
Education, n (%)	
Primary school or below	27 (5.1)
Secondary school or below	173 (33)
Trade certificate or diploma	141 (26.9)
Completed university	175 (33.3)
Language at home, n (%)	
English	363 (69.1)
Other	161 (30.7)
Socioeconomic status^a, n (%)	
IRSD ^b 1-4	123 (23.5)
IRSD 5-6	111 (21.1)
IRSD 7-8	134 (25.5)
IRSD 9-10	140 (26.6)
Private health insurance, n (%)	
Yes	249 (47.4)
No	267 (50.9)
Longstanding illness (a participant may have >1), n (%)	
No	225 (42.9)
Arthritis	115 (21.9)
Cancer	14 (2.7)
Heart disease	90 (17.1)
Diabetes	67 (12.8)
Respiratory condition	41 (7.8)
Anxiety	69 (13.1)
Depression	69 (13.1)
Other	89 (17)
Perceived health status, n (%)	
Good to excellent	400 (76.1)
Fair to poor	103 (21.5)

Characteristics	Value
Ownership of digital device (a participant may have >1), n (%)	
Computer or laptop	374 (71.2)
Mobile phone or smartphone	459 (87.4)
Tablet	241 (45.9)
Number of devices owned, mean (SD; range)	2.1 (0.9; 0-4)
Use of digital communication platform (a participant may have >1), n (%)	
Email	394 (75)
SMS text messaging	398 (75.8)
Facebook	266 (50.7)
Twitter	30 (5.7)
Instagram	104 (19.8)
Snapchat	51 (9.7)
WhatsApp or WeChat	112 (21.3)
Blogging	15 (2.9)
Forum/chat room	26 (5)
Number of platforms used, mean (SD; range)	2.7 (1.8; 0-10)
Looked for web-based information in the last 3 months, n (%)	392 (74.4)
Monitored health digitally, n (%)	183 (34.9)

^aSocioeconomic status is classified by IRSD10. This index is based on information provided by the Australian Bureau of Statistics [59]. Postcodes are divided into 10 ranks with the higher number indicating more advantaged suburbs.

^bIRSD: Index of Relative Socio-economic Disadvantage Decile 2016 of Australia.

Descriptive Statistics

The mean scale scores ranged from 2.43 (SD 0.57) for scale 7 (*Digital services that suit individual needs*) to 2.95 (SD 0.41) for scale 2 (*Understanding of health concepts and language*; Table 2). Missing values for individual items were <5%, another indication that the items were generally well understood. No floor effect was found, whereas ceiling effects were found for 8 items (range 16.2%-20.8%). These items were related to the use of technology and understanding of health knowledge, suggesting that a substantial proportion of participants were

comfortable in using technology and had good knowledge of health (Multimedia Appendix 2).

Observation during cognitive interviewing and the main survey did not identify any issue when people responded to the items for either the paper-based or web-based format. An inspection of the comments marked on the 530 completed questionnaires from the main survey found that 0.03% (15/530) of the participants put a question mark next to some items, indicating that they did not understand those items, while 0.10% (55/530) of the participants provided unclear answers. These results suggested that the items were generally understood, and the 4-point ordinal scale was acceptable.

Table 2. eHealth Literacy Questionnaire scale scores (N=525; score range 1-4).

Scale	Value, mean (SD)	Missing data
1. Using technology to process health information	2.59 (0.61)	0
2. Understanding of health concepts and language	2.95 (0.41)	0
3. Ability to actively engage with digital services	2.65 (0.68)	1
4. Feel safe and in control	2.83 (0.49)	5
5. Motivated to engage with digital services	2.63 (0.55)	0
6. Access to digital services that work	2.64 (0.45)	1
7. Digital services that suit individual needs	2.43 (0.57)	11

CTT Analysis

A range of item difficulty was found for all scales, reflecting a spectrum of difficulty levels across the relevant constructs. The scale with the smallest range of item difficulty was 7 (*Digital services that suit individual needs*; range 45%-60%). The widest range of item difficulty was observed for scale 4 (*Feel safe and in control*), ranging from 14% to 52%. Scale 7 was also the most *difficult* scale, as the difficulty level of all items was at least 45%, whereas scale 2 (*Understanding of health concepts and language*) was the *easiest* scale with 4 items <20% and the hardest item was 37% ([Multimedia Appendix 3](#)).

The chosen 1-factor Bayesian models (with informative priors for residual covariances of $df=60$) had *PPP* values that ranged from 0.19 to 0.24 for the 7 scales with all target loadings statistically significant, establishing evidence of scale homogeneity. Factor loadings were all >0.50, except for item 3 of scale 6 (*Access to digital services that work*) with a loading of 0.45. Residual variances were <0.50, except for 6 items, with item 3 recording the highest estimate of 0.80 ([Multimedia Appendix 3](#)).

A subsequent 7-factor model was fitted to the data set with 6 models tested. All models fitted the data well. The model with priors for the variance of cross-loadings set to 0.01 and inverse-Wishart df for residual covariances of 150 was chosen as the model of interest (*PPP*=.49, 95% CI for the difference between observed and replicated χ^2 values -101.40 to 108.83, *PPPP*=.92). No statistically significant cross-loadings were found for the chosen model, indicating that all items loaded only on 1 factor ([Multimedia Appendix 4](#)). Except for 4 items, all factor loadings of the chosen 7-factor model were >0.50, with item 26 of scale 2 (*Understanding of health concepts and language*) recording the lowest loading of 0.36. In addition, all cross-loadings were less than -0.20 to +0.20; that is, they could be considered approximate zero and negligible ([Table 3](#)).

Inspection of the AVE showed that the estimates of 4 scales met the first Fornell-Larcker criterion, whereas 3 scales were <0.50 (scales 2, 4, and 6). Given that these AVE estimates were based on the 7-factor model that allowed for cross-loadings and residual covariances, AVE estimates from the 1-factor models were also calculated, and the AVE estimates for the 7 scales were 0.66, 0.49, 0.72, 0.61, 0.65, 0.47, and 0.74. Hence, the AVE estimates of scales 2 and 6 were still <0.50. The second criterion of the factor's association with other factors was also not satisfied. On the basis of this criterion, only scale 2 demonstrated good discrimination with scales 4, 6, and 7, and scale 4 had good discrimination with all scales except scale 6.

Hence, there might not be sufficient discriminant validity among the scales ([Table 4](#)).

For internal consistency, the Cronbach α (range .74-.90) and composite scale reliability (range 0.73-0.90) estimates were very similar as expected. All were within the acceptable range, whereas scales 1, 3, 4, 5, and 7 had estimates >0.80, indicating good internal consistency. For test-retest, 42 participants completed the retest and ICC ranged from 0.72 to 0.95, suggesting good test-retest reliability ([Multimedia Appendix 3](#)).

The Bayesian MIMIC models for testing DIF for administration format, site area, and health setting achieved a good model fit. The model with a DIF path of 0.01 was chosen as the model of interest for the 7 scales with *PPP* ranging from .19 (scale 6) to .35 (scale 7), and all *PPPPs* were nonsignificant. No statistically significant effects of site area and health setting on the items were found. The administration format was found to have a statistically direct effect on 5 items, indicating possible DIF. However, all estimates were within the acceptable range of -0.2 to +0.2 and therefore considered negligible ([Multimedia Appendix 5](#)).

For group differences, no significant differences were found for site area and health setting, but group differences were identified for the administration format with the self-administered paper-based format scoring higher than face-to-face interviews for scales 1 (*Using technology to process health information*), 3 (*Ability to actively engage with digital services*), 5 (*Motivated to engage with digital services*), and 7 (*Digital services that suit individual needs*; [Table 5](#)). Further analysis using nonparametric tests was undertaken to explore if the 2 groups had significant differences in terms of age, education, and technology use, with a significance level set at <.05. A Mann-Whitney *U* test indicated significant difference in age for interview (median 75; $n=109$) and paper format (median 51; $n=387$; $U=7881.50$; $z=-10.00$; $P<.001$), with participants interviewed being older than those completing the self-administered paper format. A chi-square test for independence indicated a significant association between education and administration format, $\chi^2_1=0.4$ ($n=503$), $P<.001$. Among the participants being interviewed, 36.7% (40/109) did not complete secondary school, whereas only 13.2% (51/387) of participants completing the paper format did not complete secondary school. A significant difference in the number of devices was also found for interview (median 2; $n=111$) and paper format (median 2; $n=390$; $U=20,328.50$; $z=6.86$; $P<.001$), with more devices for participants using the self-administered paper format than for participants being interviewed.

Table 3. Factor loadings of the eHealth Literacy Questionnaire 7-factor Bayesian confirmatory factor analysis model with priors for cross-loadings of 0.01 and residual covariances of 150^a.

Item ^b	1. Using technology	2. Health concepts	3. Ability	4. Feel safe	5. Motivated	6. Access	7. Suit needs
1. Using technology to process health information							
Q7D1	<i>0.94^c</i>	0.02	0.02	0.00	-0.08	-0.06	-0.06
Q11D1	<i>0.89</i>	0.03	-0.01	-0.01	-0.02	-0.04	-0.08
Q13D1	<i>0.59</i>	-0.02	-0.03	0.02	0.08	0.06	0.05
Q20D1	<i>0.49</i>	-0.03	-0.01	0.02	0.05	0.06	0.08
Q25D1	<i>0.61</i>	-0.01	0.02	0.01	0.03	0.04	0.06
2. Understanding of health concepts and language							
Q5D2	0.06	<i>0.52</i>	0.03	0.00	0.04	0.01	0.01
Q12D2	0.02	<i>0.70</i>	0.01	0.02	-0.02	-0.03	-0.03
Q15D2	-0.04	<i>0.51</i>	-0.02	0.03	-0.01	0.03	0.02
Q21D2	-0.03	<i>0.67</i>	-0.01	-0.01	-0.03	-0.02	-0.02
Q26D2	0.02	<i>0.36</i>	-0.00	-0.02	0.05	0.04	0.04
3. Ability to actively engage with digital service							
Q4D3	0.00	-0.00	<i>0.68</i>	0.04	0.03	0.03	0.03
Q6D3	0.02	0.01	<i>0.88</i>	0.03	-0.02	-0.04	-0.05
Q8D3	0.03	0.02	<i>0.62</i>	0.01	0.03	0.02	0.03
Q17D3	0.00	-0.01	<i>0.88</i>	-0.02	-0.03	-0.03	-0.04
Q32D3	-0.03	0.01	<i>0.74</i>	-0.04	0.01	0.03	0.07
4. Feel safe and in control							
Q1D4	0.02	0.00	0.01	<i>0.67</i>	-0.01	-0.02	-0.03
Q10D4	0.05	0.02	0.01	<i>0.67</i>	0.04	0.01	0.00
Q14D4	0.04	0.05	0.02	<i>0.40</i>	0.05	0.05	0.03
Q22D4	-0.03	-0.02	-0.01	<i>0.86</i>	-0.03	-0.01	0.00
Q30D4	-0.01	-0.01	0.01	<i>0.74</i>	0.01	0.02	0.04
5. Motivated to engage with digital services							
Q2D5	-0.04	-0.02	-0.01	0.02	<i>0.76</i>	-0.01	-0.02
Q19D5	0.04	0.02	0.02	-0.04	<i>0.71</i>	-0.00	0.00
Q24D5	-0.02	0.01	-0.02	0.05	<i>0.67</i>	0.02	0.01
Q27D5	0.00	0.01	-0.01	0.01	<i>0.74</i>	-0.01	-0.02
Q35D5	0.04	0.00	0.03	-0.01	<i>0.72</i>	-0.00	0.00
6. Access to digital services that work							
Q3D6	-0.11	0.02	-0.05	0.06	-0.08	<i>0.59</i>	-0.08
Q9D6	0.13	-0.00	0.08	-0.03	0.05	<i>0.40</i>	0.05
Q16D6	-0.11	0.02	-0.04	0.05	-0.05	<i>0.65</i>	-0.02
Q23D6	0.05	-0.03	0.01	0.00	0.03	<i>0.61</i>	0.01
Q29D6	0.00	-0.01	-0.01	-0.01	0.02	<i>0.61</i>	-0.00
Q34D6	0.12	0.01	0.08	-0.07	0.07	<i>0.48</i>	0.07
7. Digital services that suit individual needs							
Q18D7	0.05	0.02	0.04	-0.02	-0.02	-0.03	<i>0.74</i>
Q28D7	0.00	-0.03	-0.02	0.01	0.03	0.01	<i>0.78</i>
Q31D7	-0.09	0.01	-0.05	0.07	-0.04	0.03	<i>0.85</i>

Item ^b	1. Using technology	2. Health concepts	3. Ability	4. Feel safe	5. Motivated	6. Access	7. Suit needs
Q33D7	0.02	0.00	0.04	-0.04	-0.00	-0.02	0.85

^aModel fit: posterior predictive $P=0.49$, 95% CI for the difference between observed and replicated χ^2 values -101.40 to 108.83, prior-posterior predictive $P=.92$.

^bSee [Multimedia Appendix 2](#) for truncated items.

^cItalicized values indicate statistically significant factor loadings ($P<.05$) with standardized estimates reported.

Table 4. Interfactor correlations (below diagonal), average variance extracted (diagonal), and shared variance estimates (above diagonal) for the 7 eHealth Literacy Questionnaire scales.

Scale	1. Use tech	2. Health concepts	3. Ability	4. Feel safe	5. Motivated	6. Access	7. Suit needs
1. Using technology to process health information	0.53 ^a	0.37 ^b	0.90 ^b	0.06	0.84 ^b	0.38 ^b	0.56 ^b
2. Understanding of health concepts and language	0.61	0.32 ^a	0.38 ^b	0.22	0.34 ^b	0.25	0.21
3. Ability to actively engage with digital services	0.95	0.62	0.59 ^a	0.04	0.72 ^b	0.34 ^b	0.61 ^b
4. Feel safe and in control	0.25	0.47	0.21 ^c	0.47 ^a	0.12	0.34 ^b	0.19
5. Motivated to engage with digital services	0.91	0.58	0.85	0.35	0.52 ^a	0.63 ^b	0.69 ^b
6. Access to digital services that work	0.62	0.50	0.58	0.58	0.80	0.32 ^a	0.75 ^b
7. Digital services that suit individual needs	0.75	0.46	0.78	0.43	0.83	0.87	0.65 ^a

^aThese values indicated average variance extracted by each latent variable.

^bThese values indicate that latent variable shared variance estimates exceed the average variance extracted of either or both variables.

^cStatistically not significant interfactor correlation ($P>.05$).

Table 5. Estimated effects of administration format, site area, and health setting on the 7 eHealth literacy latent variables.

Scale	Admin format ^{a,b}	Site area ^{a,c}	Health setting ^{a,d}
1. Using technology to process health information	0.38 (0.05) ^e	0.02 (0.06)	0.10 (0.06)
2. Understanding of health concepts and language	-0.02 (0.07)	-0.00 (0.07)	0.05 (0.08)
3. Ability to actively engage with digital services	0.39 (0.05)	-0.02 (0.05)	0.07 (0.06)
4. Feel safe and in control	-0.03 (0.06)	0.12 (0.06)	-0.04 (0.07)
5. Motivated to engage with digital services	0.25 (0.05)	0.03 (0.06)	0.10 (0.06)
6. Access to digital services that work	0.02 (0.06)	0.02 (0.06)	-0.02 (0.07)
7. Digital services that suit individual needs	0.22 (0.06)	-0.01 (0.06)	0.06 (0.07)

^aStandardized estimates reported; posterior SD for estimates shown in parentheses.

^bAdministration format code: 0=interview, 1=paper.

^cSite area code: 0=metropolitan, 1=regional.

^dHealth setting code: 0=private clinic, 1=community health.

^eItalicized values indicate statistically significant differences, significant if $P<.025$ (1-tailed).

IRT Analysis

The results of the 1-factor Bayesian models with all significant targeted factor loadings provided evidence of unidimensionality and local independence. Item invariance was supported by the testing of DIF for administration format, site area, and health setting. Measurement invariance across subgroups, including age, sex, education, language spoken at home, and information and communication technology use, was also established and reported by Cheng et al [20]. Hence, the results of the CTT analysis confirmed 3 of the 4 assumptions for IRT analysis. For

the final assumption of monotonicity, the test characteristic curves were examined, confirming that the probability of endorsing an item increased as the trait level increased.

Visual inspection of the item characteristic curves showed distinct peaks for the response categories along the continuum of the latent trait for the most likely responses, indicating ordered thresholds for all items ([Multimedia Appendix 6](#)). The item discrimination parameters demonstrated that items within each scale had different sensitivities to different levels of ability. All slopes of the item characteristic curves were steep, with the

steepest slope observed for item 33 of scale 7 (*Digital services that suit individual needs*), which also had the highest item discrimination parameter of 5.56. The items with lower discrimination parameters among all items were item 3 of scale 6 (*Access to digital services that work*; 0.86) and item 26 of scale 2 (*Understanding health concepts and language*; 0.88). However, the item characteristic curves of both items were still considered steep. The item location parameters also showed that items had different levels of difficulty within each scale but were not evenly distributed for scales 2 and 4. Items 15 and 21 of scale 2 (*Understanding of health concepts and language*) had very similar item location parameters of -1.19 and -1.18 , respectively. Scale 4 (*Feel safe and in control*) also had item location parameters of -0.58 for item 10 and -0.54 for item 22. The results were generally in line with the item difficulty indexes from the CTT analysis (Multimedia Appendix 3). Furthermore, the information function curves were evaluated for reliability. All items provided precise information at different levels of the latent trait, except item 14 of scale 4 (*Feel safe and in control*), which provided very low information across all levels of the trait (Multimedia Appendix 6).

Discussion

Principal Findings

This study collected and examined validity evidence based on test content, response process, and internal structure of the eHLQ in the Australian community health setting. Items and formats were easy to read and use, and items were understood as intended. The Bayesian CFA and IRT analyses confirmed the robustness of the internal structure. However, discriminant validity based on estimates of the 7-factor BSEM was not well established and will require further investigation.

The cognitive interviews were successful in identifying 1 confusing term, which was revised, and the introductory page of the questionnaire was also improved. The results combined with observation during interviews and the survey as well as the limited number of missing values provided a wealth of information in support of the validity evidence on test content and response process of the eHLQ.

The final sample size of 525 of the cross-sectional survey provided an adequate sample size for both the CTT and IRT analyses. Although the sample had more women and university-educated participants than the Australian national averages, the sociodemographic characteristics of the participants still reflected a generally diverse sample. Nevertheless, a quarter of the sample did not use email or look for web-based information, showing that people with limited use of technology or eHealth were represented in the survey. This would ensure that the validity testing results of the eHLQ were also applicable to people with potentially lower eHealth literacy. In addition, the group differences in the 4 scales identified in the administration format for paper-based and interviews further pointed to the results that these 2 groups were significantly different in terms of age, education, and technology use. As such, the purpose of an interview option as a recruitment strategy to include older people or people with lower literacy was fulfilled. By contrast, evidence of measurement invariance

across the 2 formats confirmed that responses were not influenced by interviewing bias or social desirability. Given a separate analysis of this sample found that older people also scored lower on the same 4 scales [20], the findings of the group differences between the 2 administration formats are not surprising. However, future studies may consider examining if such group differences continue to persist when the interview option is provided at random. The identified group differences also imply that the interview option should always be available such that older people or people with lower literacy are included in future eHealth literacy research to ensure that they are not being left behind in the age of digital health.

A rigorous assessment of the internal structure was undertaken using both the CTT and IRT approaches to ensure that different aspects of validity and reliability of the eHLQ data were investigated. For the CTT analysis, the Bayesian approach of applying informative priors was used. Although this modern approach may involve more steps in testing model fit, it allows for the hypothesis of approximate zeros for model parameters. Instead of being constrained to exact zeros, as in the traditional structural equation modeling approach, the approach provides a better approximation of the real world. As such, the seven 1-factor models were found to fit the data well, confirming scale homogeneity, while factor loadings and residual variances were acceptable. Estimates of internal consistency reliability were good for all scales, although scale 2 (*Understanding of health concepts and language*) and scale 6 (*Access to digital services that work*) were low but still fell within the acceptable range. Test-retest reliability was also good for all scales, indicating that the eHLQ produces stable and consistent results.

The characteristics of the test and items in accurately measuring eHealth literacy were further supported by the IRT analysis. The test and item characteristic curves demonstrated that participants with higher eHealth literacy were more likely to endorse items with agree and strongly agree. The information function curves indicated that the items could gather reliable and precise information across different levels of the underlying trait. Estimates further showed that the items had generally high sensitivity in discriminating participants with different levels of eHealth literacy. The item locations also supported the fact that the items represented different levels of difficulty. This is further verified with the item difficulty indexes from the CTT analysis, which showed the 2 estimates displaying very similar pattern, further strengthening the evidence that the items generally represented a range of difficulty levels of the latent factor. The use of MIMIC models also found no or negligible DIF for administration format, site area, and health setting, confirming measurement equivalence of the items across formats and settings.

Although it is noted that the Australian results are generally similar to the Danish validity testing results reported by Kayser et al [10], a comparison of the item location results found otherwise. The 2 results are in contradiction such that the easiest item in the Australian main sample is the hardest item in the Danish context for most of the scales. For example, in scale 1 (*Using technology to process health information*), item 7 is the easiest item for the Australian sample, but it is the hardest item for the Danish data set. As the Danish study does not report on

CTT item difficulty estimates, it cannot be used to calibrate with the Australian results. This may be due to people's different practices in using and accessing digital health between the 2 countries, as Denmark has much more regular users of web-based health systems than Australia. The real reason behind the differences is difficult to speculate and future investigation (eg, using cognitive interviewing specifically focused on the levels of response to items with contrasting difficulties in the 2 countries) may shed more light on the discrepancy.

Following the 1-factor models, a subsequent 7-factor model using informative priors for cross-loadings and residual covariances demonstrated excellent model fit of the factor structure, as hypothesized by the questionnaire developers. All target loadings were significant with acceptable factor loadings, and there was also no significant cross-loading for the chosen model of interest. Although the chosen model of interest has informative priors different from the chosen model of the Danish validity testing, the Australian data analyses generally replicate the Danish results, strengthening the evidence of the internal structure of the eHLQ.

A possible weakness in the psychometric properties of the eHLQ may be its discriminant validity. The AVE estimates suggested a lack of clear discrimination among all the scales except for scale 2 (*Understanding of health concepts and language*) and scale 4 (*Feel safe and in control*). Although AVE estimates were not investigated in the Danish validity testing, the high interfactor correlations between scales 1 and 5 and scales 6 and 7 in the Danish validity testing also suggested possible insufficient discrimination among those factors, and it was speculated that there might be some causal relationships among these scales [10]. However, the test developers argued that content analysis of the views of patients and professionals during development confirmed that these factors were indeed different constructs and decided to keep the 7 dimensions in the final model. Further investigation of discriminant validity is warranted in future validity testing of the tool.

This study provided robust validity evidence of inferences drawn from the eHLQ when used in the diverse Australian community

health settings. As this study was undertaken before the COVID-19 pandemic, which sees an increased acceptance and use of telehealth [60] as well as the widespread of misinformation and disinformation on social media [1], the eHLQ will be a useful tool for health care providers, researchers, digital health developers, and policy makers to better understand the eHealth literacy needs of individual patients and different population groups. The insights gained will help develop, implement, and evaluate digital health interventions that suit the needs of users to promote health and equity.

Limitations

A possible limitation to the validity evidence is that the sample involved only participants who spoke and understood English well. Although the eHLQ is one of the first questionnaires developed simultaneously in 2 languages to minimize cultural references, both languages are from Western culture with generally well-developed national health care systems. How the psychometric properties perform in other cultural groups and countries is not clear. Future research on the eHLQ should include validity testing in cross-cultural settings including in different contexts and use. The Danish validity testing study was undertaken in the community setting involving the general population. However, this study only included people attending community health services. Future testing of the eHLQ in other Australian settings may strengthen the validity evidence of the tool for the general population.

Conclusions

The evidence presented in this study suggests that the eHLQ is a tool with robust psychometric properties. There is support for test content, and the items are understood as intended. Although there are potential weaknesses in discriminant validity, it is reasonable to suggest that the items can provide valid and reliable assessment of the 7 constructs of eHealth literacy in the diverse Australian health settings. The eHLQ is ready to be used to identify eHealth literacy strengths and challenges and assist the development of digital health interventions to ensure that people with limited digital access and skills are not being left behind.

Acknowledgments

The authors thank Dr Mukesh Haikerwal AC, Ms Jenny Ktenidis, Ms Rori Plaza of Altona North Medical Group and Circuit Health, Ms Janine Scott, and Ms Olive Aumann of Carrington Health and Dr Ewa Piejko, Dr Adel Asaid, Dr Remon Eskander, and Dr Poate Radrekusa of St Anthony Family Medical Practice for their generous support. We also thank Prof Lisa Hanna of Deakin University for her input and Dr Polina Putrik, Visiting Fellow at Deakin University, for her assistance in data collection. RHO was funded in part through the National Health and Medical Research Council of Australia Principal Research Fellowship APP1155125.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Bayesian structural equation model of the eHealth Literacy Questionnaire with no prior. Output from Mplus.

[[DOCX File , 323 KB - jmir_v24i3e32777_app1.docx](#)]

Multimedia Appendix 2

Descriptive statistics of the eHealth Literacy Questionnaire items.

[DOCX File , 18 KB - [jmir_v24i3e32777_app2.docx](#)]

Multimedia Appendix 3

Psychometric properties of the eHealth Literacy Questionnaire single scales.

[DOCX File , 25 KB - [jmir_v24i3e32777_app3.docx](#)]

Multimedia Appendix 4

Bayesian model fit information of the eHealth Literacy Questionnaire 7-factor models.

[DOCX File , 15 KB - [jmir_v24i3e32777_app4.docx](#)]

Multimedia Appendix 5

Estimates for the direct effect of eHealth Literacy Questionnaire items on administration format, site area, and health setting.

[DOCX File , 18 KB - [jmir_v24i3e32777_app5.docx](#)]

Multimedia Appendix 6

Item characteristic curves and information function curves of the eHealth Literacy Questionnaire items (Item Response Theory for Patient-Reported Outcomes outputs).

[DOCX File , 225 KB - [jmir_v24i3e32777_app6.docx](#)]

References

1. Managing the COVID-19 infodemic: Promoting healthy behaviours and mitigating the harm from misinformation and disinformation. World Health Organization. 2020. URL: <https://tinyurl.com/4mnhb8rs> [accessed 2020-02-10]
2. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. *BMJ* 2010;341:c5814. [doi: [10.1136/bmj.c5814](https://doi.org/10.1136/bmj.c5814)]
3. Greenhalgh T, Stramer K, Bratan T, Byrne B, Russell J, Hinder S, et al. The Devil's in the detail: final report of the independent evaluation of the Summary Care Record and HealthSpace programmes. University College London, London. 2010. URL: <https://tinyurl.com/7w8y8vhr> [accessed 2022-02-22]
4. Czaja SJ, Sharit J, Lee CC, Nair SN, Hernández MA, Arana N, et al. Factors influencing use of an e-health website in a community sample of older adults. *J Am Med Inform Assoc* 2013 Mar 01;20(2):277-284 [FREE Full text] [doi: [10.1136/amiajnl-2012-000876](https://doi.org/10.1136/amiajnl-2012-000876)] [Medline: [22802269](https://pubmed.ncbi.nlm.nih.gov/22802269/)]
5. Czaja SJ, Sharit J, Nair SN. Usability of the Medicare health web site. *JAMA* 2008 Aug 20;300(7):790-792 [FREE Full text] [doi: [10.1001/jama.300.7.790-b](https://doi.org/10.1001/jama.300.7.790-b)] [Medline: [18714058](https://pubmed.ncbi.nlm.nih.gov/18714058/)]
6. 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](https://doi.org/10.2196/jmir.8.2.e9)] [Medline: [16867972](https://pubmed.ncbi.nlm.nih.gov/16867972/)]
7. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the Effectiveness of Consumer-Focused Health Information Technology Systems Through eHealth Literacy: A Framework for Understanding Users' Needs. *JMIR Hum Factors* 2015 May 20;2(1):e9 [FREE Full text] [doi: [10.2196/humanfactors.3696](https://doi.org/10.2196/humanfactors.3696)] [Medline: [27025228](https://pubmed.ncbi.nlm.nih.gov/27025228/)]
8. Monkman H, Kushniruk A. eHealth literacy issues, constructs, models, and methods for health information technology design and evaluation. *Knowledge Management & E-Learning* 2015;7(4):541-549. [doi: [10.34105/j.kmel.2015.07.036](https://doi.org/10.34105/j.kmel.2015.07.036)]
9. Norgaard O, Furstrand D, Klokke L, Karnoe A, Batterham R, Kayser L. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. *Knowl Manag E-Learn* 2015;7(4):522-540. [doi: [10.34105/j.kmel.2015.07.035](https://doi.org/10.34105/j.kmel.2015.07.035)]
10. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the eHealth literacy framework: development and initial validity testing of the eHealth literacy questionnaire (eHLQ). *J Med Internet Res* 2018 Feb 12;20(2):e36 [FREE Full text] [doi: [10.2196/jmir.8371](https://doi.org/10.2196/jmir.8371)] [Medline: [29434011](https://pubmed.ncbi.nlm.nih.gov/29434011/)]
11. Eurohealth: digital health systems. European Observatory on Health Systems and Policies. 2019. URL: <https://apps.who.int/iris/bitstream/handle/10665/326127/Eurohealth-25-2-2019-eng.pdf?sequence=7&isAllowed=y> [accessed 2022-02-22]
12. Future of digital health systems: Report on the WHO symposium on the future of digital health systems in the European Region. World Health Organization. 2019. URL: <https://apps.who.int/iris/bitstream/handle/10665/329032/9789289059992-eng.pdf> [accessed 2022-02-22]
13. McCauley D. Australians shun My Health Record with only 9 per cent ever logging in. *The Sydney Morning Herald* December. 2019. URL: <https://tinyurl.com/nfpzpuuds> [accessed 2022-02-22]
14. American Educational Research Association, American Psychological Association, National Council on Measurement in Education. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 2014.
15. Messick S. Validity. In: Linn RL, editor. *Educational Measurement*. New York: Macmillan Publishing Co, Inc; 1989:13-103.

16. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. *Am J Med* 2006 Feb;119(2):166-116. [doi: [10.1016/j.amjmed.2005.10.036](https://doi.org/10.1016/j.amjmed.2005.10.036)] [Medline: [16443422](https://pubmed.ncbi.nlm.nih.gov/16443422/)]
17. Miller K, Willson S, Chepp V, Ryan J. Analysis. In: Millier K, Chepp V, Willson S, Padilla J, editors. *Cognitive Interviewing Methodology*. Somerset, United States: John Wiley & Sons, Incorporated; 2014:35-50.
18. Rios J, Wells C. Validity evidence based on internal structure. *Psicothema* 2014;26(1):108-116. [doi: [10.7334/psicothema2013.260](https://doi.org/10.7334/psicothema2013.260)] [Medline: [24444738](https://pubmed.ncbi.nlm.nih.gov/24444738/)]
19. Zumbo B, Chan E. *Validity and Validation in Social, Behavioral, and Health Sciences*. Switzerland: Springer International Publishing; 2014.
20. Cheng C, Elsworth G, Osborne RH. Validity evidence based on relations to other variables of the ehealth literacy questionnaire (eHLQ): Bayesian approach to test for known-groups validity. *J Med Internet Res* 2021 Oct 14;23(10):e30243 [FREE Full text] [doi: [10.2196/30243](https://doi.org/10.2196/30243)] [Medline: [34647897](https://pubmed.ncbi.nlm.nih.gov/34647897/)]
21. Bailey C, Tully M, Cooke J. An investigation into the content validity of the Antimicrobial Self-Assessment Toolkit for NHS Trusts (ASAT v15a) using cognitive interviews with antimicrobial pharmacists. *J Clin Pharm Ther* 2015 Apr 10;40(2):208-212. [doi: [10.1111/jcpt.12248](https://doi.org/10.1111/jcpt.12248)] [Medline: [25678341](https://pubmed.ncbi.nlm.nih.gov/25678341/)]
22. Bobrovitz N, Santana MJ, Kline T, Kortbeek J, Stelfox HT. The use of cognitive interviews to revise the Quality of Trauma Care Patient-Reported Experience Measure (QTAC-PREM). *Qual Life Res* 2015 Aug 15;24(8):1911-1919. [doi: [10.1007/s11136-015-0919-5](https://doi.org/10.1007/s11136-015-0919-5)] [Medline: [25589232](https://pubmed.ncbi.nlm.nih.gov/25589232/)]
23. Willis G. *Cognitive interviewing: a "How to" guide*. Research Training Institute. 1999. URL: <https://123docz.net/document/1098481-cognitive-interviewing-a-how-to-guide-pptx.htm> [accessed 2022-02-22]
24. Beatty PC, Willis GB. Research synthesis: the practice of cognitive interviewing. *Public Opin Q* 2007 Jun 05;71(2):287-311. [doi: [10.1093/poq/nfm006](https://doi.org/10.1093/poq/nfm006)] [Medline: [25977055](https://pubmed.ncbi.nlm.nih.gov/25977055/)]
25. Benítez I, Padilla J. Analysis of nonequivalent assessments across different linguistic groups using a mixed methods approach. *J Mixed Methods Res* 2013 May 13;8(1):52-68 [FREE Full text] [doi: [10.1177/1558689813488245](https://doi.org/10.1177/1558689813488245)]
26. DeWalt DA, Rothrock N, Yount S, Stone AA, PROMIS Cooperative Group. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care* 2007 May;45(5 Suppl 1):12-21 [FREE Full text] [doi: [10.1097/01.mlr.0000254567.79743.e2](https://doi.org/10.1097/01.mlr.0000254567.79743.e2)] [Medline: [17443114](https://pubmed.ncbi.nlm.nih.gov/17443114/)]
27. Solorio R, Ayala NC, Paez E, Skalicky AM, Morales LS. Use of cognitive interviews to adapt PROMIS measurement items for Spanish speakers living with HIV. *AIDS Res Treat* 2016;2016:8340863 [FREE Full text] [doi: [10.1155/2016/8340863](https://doi.org/10.1155/2016/8340863)] [Medline: [27022480](https://pubmed.ncbi.nlm.nih.gov/27022480/)]
28. DeVellis RF. *Classical test theory*. *Med Care* 2006 Nov;44(11 Suppl 3):50-59 [FREE Full text] [doi: [10.1097/01.mlr.0000245426.10853.30](https://doi.org/10.1097/01.mlr.0000245426.10853.30)] [Medline: [17060836](https://pubmed.ncbi.nlm.nih.gov/17060836/)]
29. Raykov T, Marcoulides GA. *Introduction to Psychometric Theory*. New York: Routledge; 2010.
30. *IBM SPSS Statistics for Mac, Version 27*. Armonk, NY: IBM Corp; 2020.
31. Muthén L, Muthén B. *Mplus (Mac) Version 8*. Los Angeles, CA: Muthén & Muthén; 2017.
32. Cai L, Thissen D, du Toli S. *IRTPRO 4.20 for Windows*. Chapel Hill, NC: Vector Psychometric Group; 2019.
33. Fayers P, Machin D. *Quality of Life : The Assessment, Analysis and Interpretation of Patient-reported Outcomes*. West Sussex: John Wiley & Sons; 2007.
34. Katajapuu N, Laimi K, Heinonen A, Saltychev M. Floor and ceiling effects of the World Health Organization Disability Assessment Schedule 2.0 among patients with chronic musculoskeletal pain. *Int J Rehabil Res* 2019 Jun;42(2):190-192. [doi: [10.1097/MRR.000000000000339](https://doi.org/10.1097/MRR.000000000000339)] [Medline: [30762604](https://pubmed.ncbi.nlm.nih.gov/30762604/)]
35. Lim CR, Harris K, Dawson J, Beard DJ, Fitzpatrick R, Price AJ. Floor and ceiling effects in the OHS: an analysis of the NHS PROMs data set. *BMJ Open* 2015 Jul 27;5(7):e007765 [FREE Full text] [doi: [10.1136/bmjopen-2015-007765](https://doi.org/10.1136/bmjopen-2015-007765)] [Medline: [26216152](https://pubmed.ncbi.nlm.nih.gov/26216152/)]
36. Nolte S. *Approaches to the Measurement of Outcomes of Chronic Disease Self-management Interventions Using a Self-report Inventory*. Melbourne, Australia: RMIT University; 2008. URL: <https://researchrepository.rmit.edu.au/esploro/outputs/doctoral/Approaches-to-the-measurement-of-outcomes/9921861708201341> [accessed 2022-02-22]
37. Yuan KH, Bentler PM. Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. *Sociolog Methodol* 2000;30(1):165-200. [doi: [10.1111/0081-1750.00078](https://doi.org/10.1111/0081-1750.00078)]
38. Raykov T, Marcoulides GA. On the relationship between classical test theory and item response theory: from one to the other and back. *Educ Psychol Meas* 2016 Apr;76(2):325-338 [FREE Full text] [doi: [10.1177/0013164415576958](https://doi.org/10.1177/0013164415576958)] [Medline: [29795868](https://pubmed.ncbi.nlm.nih.gov/29795868/)]
39. Osborne RH, Batterham RW, Elsworth GR, Hawkins M, Buchbinder R. The grounded psychometric development and initial validation of the Health Literacy Questionnaire (HLQ). *BMC Public Health* 2013;13:658 [FREE Full text] [doi: [10.1186/1471-2458-13-658](https://doi.org/10.1186/1471-2458-13-658)] [Medline: [23855504](https://pubmed.ncbi.nlm.nih.gov/23855504/)]
40. Raykov T. Estimation of composite reliability for congeneric measures. *Appl Psychol Measur* 2016 Jul 27;21(2):173-184. [doi: [10.1177/01466216970212006](https://doi.org/10.1177/01466216970212006)]
41. 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](https://doi.org/10.5116/ijme.4dfb.8dfd)] [Medline: [28029643](https://pubmed.ncbi.nlm.nih.gov/28029643/)]

42. Cappelleri JC, Lundy JJ, Hays RD. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin Ther* 2014 May;36(5):648-662 [FREE Full text] [doi: [10.1016/j.clinthera.2014.04.006](https://doi.org/10.1016/j.clinthera.2014.04.006)] [Medline: [24811753](https://pubmed.ncbi.nlm.nih.gov/24811753/)]
43. Bujang MA, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. *Archiv Orofac Sci* 2017;12(1):1-11 [FREE Full text]
44. McCoach DB, Gable RK, Madura JP. Evidence based on the internal structure of the instrument: factor analysis. In: *Instrument Development in the Affective Domain*. New York, NY: Springer; 2013:109-161.
45. Muthén B, Asparouhov T. Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychol Methods* 2012 Sep;17(3):313-335. [doi: [10.1037/a0026802](https://doi.org/10.1037/a0026802)] [Medline: [22962886](https://pubmed.ncbi.nlm.nih.gov/22962886/)]
46. Asparouhov T, Muthén B, Morin AJ. Bayesian structural equation modeling with cross-loadings and residual covariances. *J Manag* 2015 Jun 30;41(6):1561-1577. [doi: [10.1177/0149206315591075](https://doi.org/10.1177/0149206315591075)]
47. Hoijtink H, van de Schoot R. Testing small variance priors using prior-posterior predictive P values. *Psychol Methods* 2018 Sep;23(3):561-569. [doi: [10.1037/met0000131](https://doi.org/10.1037/met0000131)] [Medline: [28368177](https://pubmed.ncbi.nlm.nih.gov/28368177/)]
48. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Mark Res* 1981 Feb;18(1):39-50. [doi: [10.2307/3151312](https://doi.org/10.2307/3151312)]
49. Farrell AM. Insufficient discriminant validity: a comment on Bove, Pervan, Beatty, and Shiu (2009). *J Bus Res* 2010 Mar;63(3):324-327. [doi: [10.1016/j.jbusres.2009.05.003](https://doi.org/10.1016/j.jbusres.2009.05.003)]
50. Meade AW, Lautenschlager GJ. A comparison of item response theory and confirmatory factor analytic methodologies for establishing measurement equivalence/invariance. *Organ Res Methods* 2004 Oct 1;7(4):361-388. [doi: [10.1177/1094428104268027](https://doi.org/10.1177/1094428104268027)]
51. Osterlind SJ, Everson HT. *Differential Item Functioning*. Thousand Oaks: Sage Publication; 2009.
52. Gallo JJ, Anthony JC, Muthén BO. Age differences in the symptoms of depression: a latent trait analysis. *J Gerontol* 1994 Nov;49(6):251-264. [doi: [10.1093/geronj/49.6.p251](https://doi.org/10.1093/geronj/49.6.p251)] [Medline: [7963280](https://pubmed.ncbi.nlm.nih.gov/7963280/)]
53. Asparouhov T, Muthén B. Prior-posterior predictive p-values. *Mplus Web Notes*: No. 22. 2017. URL: <https://www.statmodel.com/download/PPPP.pdf> [accessed 2022-02-22]
54. Nguyen TH, Han H, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. *Patient* 2014 Jan 9;7(1):23-35 [FREE Full text] [doi: [10.1007/s40271-013-0041-0](https://doi.org/10.1007/s40271-013-0041-0)] [Medline: [24403095](https://pubmed.ncbi.nlm.nih.gov/24403095/)]
55. Baker F. *The Basics of Item Response Theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation; 2001.
56. Ong CW, Pierce BG, Woods DW, Twohig MP, Levin ME. The Acceptance and Action Questionnaire – II: an Item Response Theory Analysis. *J Psychopathol Behav Assess* 2018 Aug 16;41(1):123-134. [doi: [10.1007/s10862-018-9694-2](https://doi.org/10.1007/s10862-018-9694-2)]
57. Edelen MO, Reeve BB. Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Qual Life Res* 2007;16 Suppl 1:5-18. [doi: [10.1007/s11136-007-9198-0](https://doi.org/10.1007/s11136-007-9198-0)] [Medline: [17375372](https://pubmed.ncbi.nlm.nih.gov/17375372/)]
58. Toland MD. Practical guide to conducting an item response theory analysis. *J Early Adolesc* 2013 Nov 19;34(1):120-151. [doi: [10.1177/0272431613511332](https://doi.org/10.1177/0272431613511332)]
59. 2033.0.55.001 - Census of Population and Housing: Socio-Economic Indexes for Areas (SEIFA), Australia, 2016. Australian Bureau of Statistics. 2018. URL: <http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/2033.0.55.0012016?OpenDocument> [accessed 2022-02-22]
60. Jahns R, Badmaeva T, Bitzer M. Impact assessment survey. How Corona impacts the global digital health industry. *Research2guidance*. 2020 Apr 14. URL: <https://tinyurl.com/48ydyxkp> [accessed 2022-02-22]

Abbreviations

- AVE**: average variance extracted
- BSEM**: Bayesian structural equation modeling
- CFA**: confirmatory factor analysis
- CTT**: classical test theory
- DIF**: differential item functioning
- eHLQ**: eHealth Literacy Questionnaire
- ICC**: intraclass correlation coefficient
- IRT**: item response theory
- MIMIC**: multiple-indicators, multiple-causes
- PPP**: posterior predictive P
- PPPP**: prior-posterior predictive P

Edited by R Kukafka; submitted 10.08.21; peer-reviewed by B Raffaelli, RM Payo, J Lee, S Weiwei; comments to author 19.10.21; revised version received 08.11.21; accepted 03.12.21; published 08.03.22.

Please cite as:

Cheng C, Elsworth GR, Osborne RH

Validity Evidence of the eHealth Literacy Questionnaire (eHLQ) Part 2: Mixed Methods Approach to Evaluate Test Content, Response Process, and Internal Structure in the Australian Community Health Setting

J Med Internet Res 2022;24(3):e32777

URL: <https://www.jmir.org/2022/3/e32777>

doi: [10.2196/32777](https://doi.org/10.2196/32777)

PMID: [35258475](https://pubmed.ncbi.nlm.nih.gov/35258475/)

©Christina Cheng, Gerald R Elsworth, Richard H Osborne. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 08.03.2022. 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.

Original Paper

Comparing SF-36 Scores Collected Through Web-Based Questionnaire Self-completions and Telephone Interviews: An Ancillary Study of the SENTIPAT Multicenter Randomized Controlled Trial

Ayşe Açma¹, MPH; Fabrice Carrat², MD, PhD; Gilles Hejblum¹, PhD[‡]

¹Sorbonne Université, INSERM, Institut Pierre Louis d'Épidémiologie et de Santé Publique, Paris, France

²Sorbonne Université, INSERM, Institut Pierre Louis d'Épidémiologie et de Santé Publique, AP-HP, Hôpital Saint-Antoine, Unité de Santé Publique, Paris, France

[‡]on behalf of the SENTIPAT study group

Corresponding Author:

Gilles Hejblum, PhD

Sorbonne Université

INSERM

Institut Pierre Louis d'Épidémiologie et de Santé Publique

27 Rue Chaligny

Paris, 75012

France

Email: gilles.hejblum@inserm.fr

Abstract

Background: The 36-Item Short Form Health Survey (SF-36) is a popular questionnaire for measuring the self-perception of quality of life in a given population of interest. Processing the answers of a participant comprises the calculation of 10 scores corresponding to 8 scales measuring several aspects of perceived health and 2 summary components (physical and mental). Surprisingly, no study has compared score values issued from a telephone interview versus those from an internet-based questionnaire self-completion.

Objective: This study aims to compare the SF-36 score values issued from a telephone interview versus those from an internet-based questionnaire self-completion.

Methods: Patients with an internet connection and returning home after hospital discharge were enrolled in the SENTIPAT multicenter randomized trial on the day of discharge. They were randomized to either self-completing a set of questionnaires using a dedicated website (internet group) or providing answers to the same questionnaires administered during a telephone interview (telephone group). This ancillary study of the trial compared SF-36 data related to the posthospitalization period in these 2 groups. To anticipate the potential unbalanced characteristics of the responders in the 2 groups, the impact of the mode of administration of the questionnaire on score differences was investigated using a matched sample of individuals originating from the internet and telephone groups (1:1 ratio), in which the matching procedure was based on a propensity score approach. SF-36 scores observed in the internet and telephone groups were compared using the Wilcoxon-Mann-Whitney test, and the score differences between the 2 groups were also examined according to Cohen effect size.

Results: Overall, 29.2% (245/840) and 75% (630/840) of SF-36 questionnaires were completed in the internet and telephone groups, respectively ($P < .001$). Globally, the score differences between groups before matching were similar to those observed in the matched sample. Mean scores observed in the telephone group were all above the corresponding values observed in the internet group. After matching, score differences in 6 out of the 8 SF-36 scales were statistically significant, with a mean difference greater than 5 for 4 scales and an associated mild effect size ranging from 0.22 to 0.29, and with a mean difference near this threshold for 2 other scales (4.57 and 4.56) and a low corresponding effect size (0.18 and 0.16, respectively).

Conclusions: The telephone mode of administration of SF-36 involved an interviewer effect, increasing SF-36 scores. Questionnaire self-completion via the internet should be preferred, and surveys combining various administration methods should be avoided.

Trial Registration: ClinicalTrials.gov NCT01769261; <https://www.clinicaltrials.gov/ct2/show/record/NCT01769261>

KEYWORDS

Bias, Epidemiologic; Effect Modifier, Epidemiologic; Forms as Topic; Interviews, Telephone; Internet; Patient Reported Outcome Measures; Patient Satisfaction; Quality of Life; Surveys and Questionnaires

Introduction

The 36-Item Short Form Health Survey (SF-36) is a popular questionnaire for measuring the self-perception of quality of life (QoL) in a given population of interest [1-3]: a query exploring the presence of the term *SF-36* in the title or the abstract of PubMed records retrieved 22,184 documents on September 28, 2021. SF-36 has been made available in 50 different languages, including French [4]. Although the SF-36 was initially developed as a paper-pencil format auto-questionnaire, the use of telephone interviews has also been reported for collecting SF-36 data [5-8]. Self-completion via the internet has been reported as a validated administration mode by Bell and Kahn [9] in 1996, and since then, with the spread of the internet and computers, several other computerized or internet-based formats have been applied in different studies [10-12].

Several randomized trials compared the SF-36 scores issued from different administration modes, such as paper versus the internet [13-17] or telephone versus paper [18-26]. Telephone interview is a common mode of questionnaire administration for several reasons, including the potential to increase response rate [24-26], practical convenience if other data of the study are already being collected via telephone, and exploring QoL in some special populations such as older patients. On the other hand, self-completion via the internet has advantages such as avoiding any potential response bias related to the interviewer effect [18], being potentially a simpler organization for collecting SF-36 data, and being associated with lower costs. However, and surprisingly, to our knowledge, no study has compared telephone interview and internet-based auto-questionnaire methods for collecting SF-36 data to investigate whether they can be used as alternative methods in mixed mode data collection procedures according to participant preferences and minimize the possible selection bias. This study investigated such questions in detail, owing to the availability of SF-36 data that had been collected in a multicenter randomized trial.

The SENTIPAT trial [27] explored the concept of sentinel patients who would voluntarily report their health evolution on a dedicated website. Participants enrolled in this trial were randomized to either the internet or the telephone group; patients in the internet group were invited to self-complete questionnaires on their health evolution after their hospital discharge via a dedicated website, whereas patients in the telephone group were invited to complete the same questionnaires through telephone interviews. However, 2 previous studies issued from the SENTIPAT trial have been reported: the first introduced an original questionnaire developed in the SENTIPAT study to investigate the opinion of patients about the organization of their hospital discharge [28], and the second introduced the I-Satis questionnaire, a questionnaire that was distributed in

hospitals at a national level in France to investigate patient satisfaction at the time of the SENTIPAT trial [29]. As the SF-36 questionnaire was also included in the SENTIPAT trial, the corresponding collected data were a perfect opportunity to precisely investigate the influence of the mode of administration of the questionnaire on SF-36 scores. This investigation is the aim of the ancillary study of the SENTIPAT trial reported here.

Methods

This research was an ancillary study of the multicenter, randomized SENTIPAT trial [27]. We took advantage of the trial to investigate the impact of the mode of administration of the SF-36 questionnaire on SF-36 scores.

Population

Briefly, as previously reported [28,29], participants recruited consecutively from 5 different volunteer units (hepatogastroenterology, gastrointestinal enterology and nutrition, general and digestive surgery, infectious and tropical diseases, and internal medicine) of the Hôpital Saint-Antoine were enrolled in the SENTIPAT trial. Patients with internet access at home, aged ≥ 18 years, not cognitively impaired and without a behavioral disorder, speaking French, returning home after hospitalization, and not opposed to participating in the trial were eligible for inclusion.

Inpatients were enrolled on the day of hospital discharge by a clinical research technician of the trial. At that time, the patients were informed about the study. Eligible patients not opposed to participating in the study were randomized into two parallel groups—internet or telephone follow-up (inherently resulting in an open-label trial)—in a ratio of 1:1. On the basis of centralized randomization that allocated the eligible patient either to the internet or to the telephone group through a website and using an underlying permutation block randomization stratified by hospital unit, a computer-generated list of permutations was established by a statistician independent of the study. At the time of patient inclusion, the technician also collected baseline variables (length of stay, sex, age, relationship status, level of education, activity, and type of insurance). The patient was then informed and discharged with documents explaining the corresponding questionnaire administration. A total of 1680 eligible patients (840 randomized in the internet group and telephone group each) were enrolled in the SENTIPAT trial between February 25, 2013, and September 8, 2014.

Ethics Approval

The SENTIPAT study was approved by the Comité de Protection des Personnes Île de France IX (decision CPP-IDF IX 12-014; June 12, 2012), the Comité Consultatif sur le Traitement de l'Information en matière de Recherche dans le domaine de la Santé (decision 12.365; June 20, 2012), and the

Commission Nationale de l'Informatique et des Libertés (decision DR-2012-582; December 12, 2012). According to the French law in force at the time of the study, the formal consent of participants was waived and replaced by the following: patients received full information on their participation in the study, and the nonopposition of each participant in the study was notified (including date of nonopposition declaration) in the SENTIPAT study register.

Survey Administration

Patients in the internet group had access to the French version of the SF-36 questionnaire 40 days after discharge on a website dedicated to SENTIPAT. Oral and written instructions had been delivered to these patients for a personal connection to the SENTIPAT website, and they received 1 reminder email per week for 3 weeks in case of nonresponse. Patients in the telephone group were interviewed by telephone approximately 42 days after discharge, and the data were simultaneously entered into the system by the interviewer using a website interface identical to that used in the internet group. The appointments for the telephone interviews of the patients in the telephone group were scheduled at the moment of patient inclusion, and up to 3 calls were tried whenever the first call did not reach the patient.

SF-36 Questionnaire and Score Calculations

The 8 scale scores and the 2 summary scores of SF-36 were calculated according to the Medical Outcomes Study SF-36 French scoring manual [30]. The main lines of the corresponding process can be summarized as follows. The SF-36 questionnaire is composed of 36 items. Completion of the SF-36 questionnaire consists in choosing one of the proposed precoded answers for each of the 36 items in the questionnaire. The analysis of 35 items (an item that relates to the evolution of perceived health is not involved in any score calculation) comprises a structured calculation of 10 scores corresponding to 8 scales measuring several aspects of perceived health and 2 summary components. The eight scales and the corresponding number of questionnaire items involved are as follows: physical functioning (10 items), role-physical (RP; corresponding to role limitations because of physical problems; 4 items), bodily pain (3 items), general health (5 items), vitality (4 items), social functioning (2 items), role-emotional (corresponding to role limitations because of emotional problems; 3 items), and mental health (5 items). The raw score of each scale was computed by the algebraic sum of the corresponding item values (the values assigned to each precoded answer were calibrated) and then normalized to a score value ranging from 0 (lowest possible) to 100 (highest possible). According to the recommendations, scale score calculations were performed only if at least half of the items involved were answered, and in such a case, missing item data were treated with a person-specific approach that uses the average score of the completed items on the same scale. Finally, the two remaining scores, the physical and mental component summary scores, were obtained by assigning predefined specific weights to each of the 8 scale scores.

Statistical Analyses

Bivariate analyses were performed using Fisher exact test or chi-square test of independence for categorical variables and the Wilcoxon-Mann-Whitney test for quantitative variables. The latter test was notably used to compare the SF-36 score differences between the internet and telephone groups. Several authors have discussed the task of interpreting observed differences in terms of *clinically meaningful* differences [31-33]. In this study, in addition to the abovementioned statistical test, the differences in SF-36 scores between the internet and telephone groups were also examined using two popular approaches: on the one hand, the effect size of the difference was considered according to Cohen's effect size index with corresponding small, medium, and large values at 0.2, 0.5, and 0.8, respectively [34]; on the other hand, we considered a threshold difference of 5 points, as proposed by Ware et al [33] for defining a clinically and socially relevant difference between 2 compared scores. The internal reliability of the SF-36 was evaluated by Cronbach's α coefficient calculation for the 8 scales and was considered acceptable if the α value was $>.7$. All statistical analyses of the study were performed using R freeware (version 3.3.3; R Foundation for Statistical Computing).

The difference between the observed SF-36 score estimates in responders of the internet group and responders of the telephone group may be mainly due to two features: (1) the difference in the mode of administration of the questionnaire, strictly speaking (self-completion of the patient via the internet vs completion of a research technician via a telephone interview with the patient), and (2) unbalanced characteristics of the individuals in the 2 groups issued from a selection bias of the responders (an unavoidable situation inherent to the modes of administration of the questionnaire). Assessing the respective impact of these 2 features on the observed differences between the SF-36 scores observed in internet and telephone responders is of primary importance, and to get more insight into this issue, we developed a procedure in which responders of the internet group were matched to similar responders of the telephone group according to their baseline characteristics, and we further examined how the score differences between the 2 groups changed in this matched sample, as compared with the score differences observed in the initial unmatched populations.

Internet responders were matched to telephone responders according to a propensity score-based procedure, and the R package MatchIt [35] was used to match each internet responder to the nearest telephone responder in a 1:1 ratio. The following baseline variables were included in the logistic regression model of the propensity score (propensity for being an internet responder vs being a telephone responder) as independent variables: age, length of stay, education, employment (unemployed because of health, retired or unemployed, job seeker, employed, and student), income, relationship status, and type of health insurance. We also forced each pair to be strictly identical according to three additional qualitative variables also included in the logistic regression model as independent variables: sex (male or female), type of hospitalization (conventional, weekly, or day-care hospitalization), and hospital ward (general and digestive surgery, gastroenterology and

nutrition, hepatogastroenterology, infectious and tropical diseases, or internal medicine).

yielded the enrollment of 840 participants to the SENTIPAT study in each arm. Table 1 shows the baseline characteristics of the patients who constituted the population investigated in this study in each group and according to the responder or nonresponder status of the participants.

Results

Enrollment of the Participants in the SENTIPAT Study

Figure 1 presents the flowchart of the study. The randomization of the participants in either the internet or telephone group

Figure 1. Flow of participants through the study. SF-36: 36-Item Short Form Health Survey.

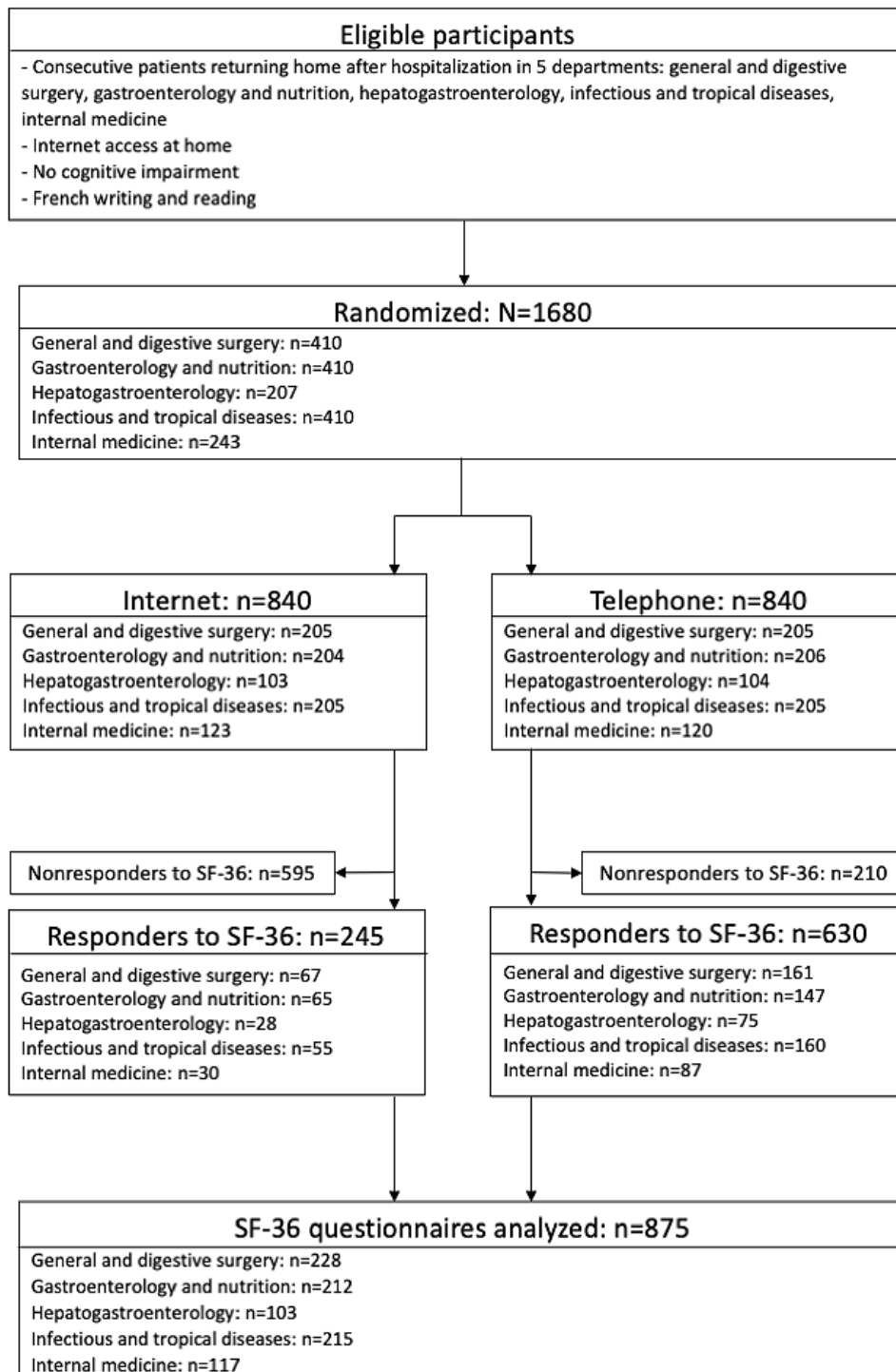


Table 1. Demographic characteristics of responders and nonresponders in the internet and telephone groups (N=1640).

Feature	Internet		Telephone	
	Responders (n=245)	Nonresponders (n=595)	Responders (n=630)	Nonresponders (n=210)
Sex, n (%)				
Female	109 (44.5)	269 (45.2)	254 (40.3)	103 (49)
Male	136 (55.5)	326 (54.8)	376 (59.7)	107 (51)
Age (years)				
Values, mean	49.5	46.6	47.2	43.8
Values, median (IQR)	50 (37-61)	47 (33-59)	47 (34-58)	41 (30-54)
Length of stay (days)				
Values, mean	4.0	4.0	4.0	4.1
Values, median (IQR)	1 (1-5)	1 (1-5)	1 (1-5)	1 (1-6)
Type of hospitalization, n (%)				
Conventional	102 (41.6)	256 (43)	269 (42.7)	91 (43.3)
1-day stay	120 (49)	285 (47.9)	297 (47.1)	103 (49.1)
Week stay	23 (9.4)	54 (9.1)	64 (10.2)	16 (7.6)
Ward, n (%)				
General and digestive surgery	67 (27.3)	138 (23.2)	161 (25.6)	44 (21)
Gastroenterology and nutrition	65 (26.5)	139 (23.4)	147 (23.3)	59 (28.1)
Hepatogastroenterology	28 (11.4)	75 (12.6)	75 (11.9)	29 (13.8)
Infectious and tropical diseases	55 (22.4)	150 (25.2)	160 (25.4)	45 (21.4)
Internal medicine	30 (12.2)	93 (15.6)	87 (13.8)	33 (15.7)
Employment, n (%)				
Currently employed	158 (65)	353 (59.3)	375 (59.5)	132 (63.2)
Job seeker	17 (7)	43 (7.2)	47 (7.5)	15 (7.2)
Retired	47 (19.3)	98 (16.5)	101 (16)	29 (13.9)
Student	6 (2.5)	38 (6.4)	48 (7.6)	17 (8.1)
Does not work because of health	11 (4.5)	48 (8.1)	49 (7.8)	11 (5.3)
Without work	2 (0.8)	9 (1.5)	8 (1.3)	4 (1.9)
Other	2 (0.8)	6 (1)	2 (0.3)	1 (0.5)
Type of employment, n (%)				
Farmer	0 (0)	1 (0)	0 (0)	0 (0)
Self-employed or trader	4 (1.6)	25 (4.2)	27 (4.3)	11 (5.3)
Manager	80 (32.7)	135 (22.7)	159 (25.2)	49 (23.4)
Intermediate profession	39 (15.9)	91 (15.3)	105 (16.7)	31 (14.8)
Middle-class occupation	52 (21.2)	135 (22.7)	123 (19.5)	55 (26.3)
Employee	5 (2)	20 (3.4)	25 (4)	8 (3.8)
Worker	42 (17.1)	83 (13.9)	92 (14.6)	22 (10.5)
No work	23 (9.4)	105 (17.6)	99 (15.7)	33 (15.8)
Level of education, n (%)				
Primary or less	18 (7.3)	58 (9.7)	47 (7.5)	31 (14.8)
High school	75 (30.6)	193 (32.4)	178 (28.3)	60 (28.7)
Superior short time	37 (15.1)	95 (16)	94 (14.9)	33 (15.8)

Feature	Internet		Telephone	
	Responders (n=245)	Nonresponders (n=595)	Responders (n=630)	Nonresponders (n=210)
Graduate or postgraduate	115 (46.9)	249 (41.8)	311 (49.4)	85 (40.7)
Relationship status, n (%)				
Living alone ^a	103 (42)	291 (48.9)	293 (46.5)	121 (57.9)
Living as a couple ^b	142 (58)	304 (51.1)	337 (53.5)	88 (42.1)
Income level (€),^c n (%)				
<450	6 (2.4)	28 (4.7)	31 (4.9)	10 (4.8)
450-1000	3 (1.2)	37 (6.2)	31 (4.9)	11 (5.3)
1000-1500	17 (6.9)	61 (10.3)	51 (8.1)	17 (8.1)
1500-2100	34 (13.9)	75 (12.6)	78 (12.4)	27 (12.9)
2100-2800	26 (10.6)	70 (11.8)	66 (10.5)	25 (12)
2800-4200	44 (18)	79 (13.3)	108 (17.1)	28 (13.4)
≥4200	43 (17.6)	64 (10.8)	82 (13)	16 (7.7)
No response	72 (29.4)	181 (30.4)	183 (29)	75 (35.9)
Type of insurance, n (%)				
State medical help or universal health insurance	2 (0.8)	26 (4.4)	24 (3.8)	8 (3.8)
Compulsory health insurance	15 (6.1)	43 (7.2)	43 (6.8)	26 (12.4)
Compulsory health insurance plus complementary private health insurance	228 (93.1)	526 (88.4)	563 (89.4)	175 (83.7)

^aSingle, widowed, divorced, or separated.

^bMarried, living together under a civil solidarity pact, or simply living together without legal ties.

^c€ (in 2013)=US \$0.71 (in 2022).

Response Rate, Delay of Questionnaire Completion, and Internal Validity of Questionnaire Completion

The response rate observed in the intervention group (245/840, 29.2%) was significantly lower ($P<.001$) than that observed in the telephone group (630/840, 75%). The median (IQR) delay between hospital discharge and questionnaire completion was 42 (40-46) and 42 (42-46) days in responders of the internet and telephone groups, respectively.

In terms of internal validity of questionnaire completion, Cronbach's α values calculated for each of the 8 scales comprising the SF-36 form in the internet and telephone groups ([Multimedia Appendix 1](#)) were all $>.7$, which is the threshold value considered as acceptable.

Assessment of the Procedure Matching the Responders of the Internet Group With Responders of the Telephone Group

The matching procedure matched the 245 responders in the internet group (no individual was dropped) with 245 individuals in the telephone group. The standardized mean difference of the global distance between internet and telephone groups was 0.4167 and 0.0215 before and after matching, respectively, with a corresponding balance improvement of 95%. [Figure 2](#) details the standardized mean differences between the internet and telephone groups observed on baseline variables before and after the matching procedure. The differences between the internet and telephone groups before matching were globally dramatically decreased after matching, indicating that the matching procedure successfully yielded two populations, internet and telephone, which were highly comparable in terms of the baseline variables.

Figure 2. Differences in baseline variables between the internet and telephone responders before and after the matching procedure.



SF-36 Score Differences According to the Mode of Administration of the Questionnaire

Figure 3 shows the differences between the internet and telephone groups, before and after matching, for the 8 scales and the 2 summary measures composing SF-36. Figure 3 indicates that the matching procedure had a limited impact on

the differences observed between the internet and telephone groups in each of the components of SF-36; regardless of the value of the difference before matching, the corresponding difference after matching appeared similar. Importantly, the means observed in the telephone group were all above the corresponding values observed in the internet group.

Figure 3. Observed mean score differences (telephone–internet) of SF-36 scales and summary components before and after matching. SF-36: 36-Item Short Form Health Survey.

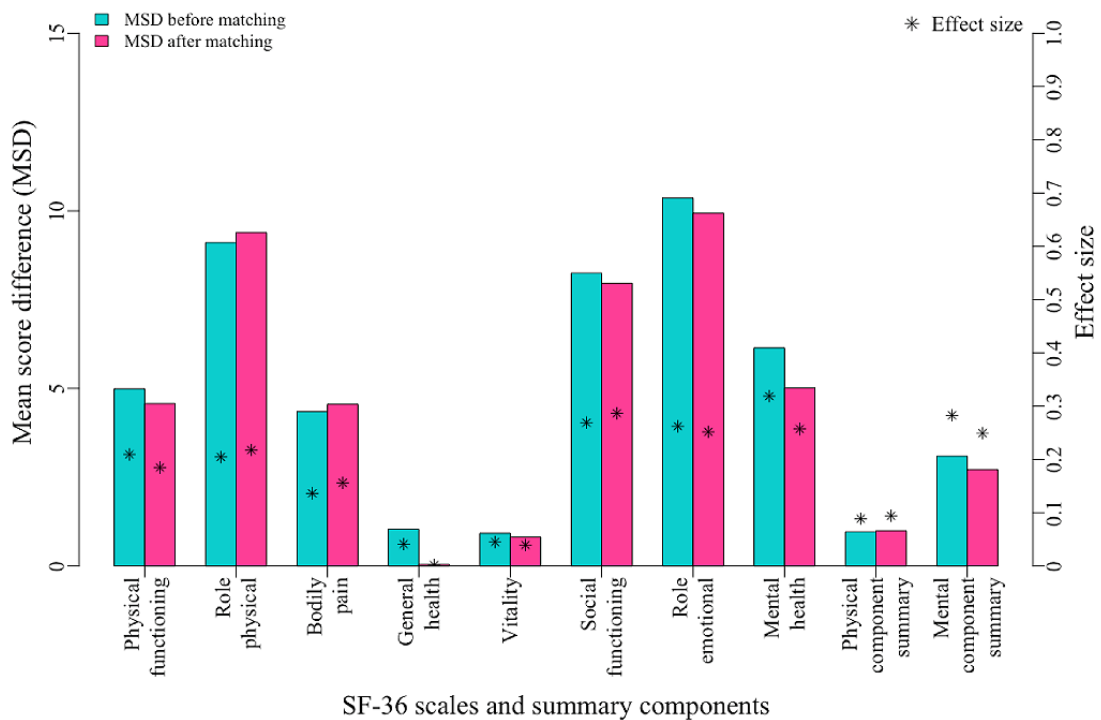


Table 2 details the results observed after matching. The mean difference between the internet and telephone groups was >5

(threshold recommended for declaring that the difference corresponds to a significant clinical status) for four scales (RP,

social functioning, role-emotional, and mental health) with an associated effect size ranging from 0.22 to 0.29, close to the value of 0.2, which is defined as a small effect size by Cohen. Moreover, the difference approached this threshold for 2 other scales (4.57 and 4.56 for physical functioning and bodily pain, respectively), with small corresponding effect size, 0.18 and 0.16, respectively. The abovementioned 6 differences were all

statistically significant (Table 2). In contrast, the observed mean difference between telephone and internet was low for the remaining 2 scales (0.04 and 0.82 for general health and vitality, respectively) and not significant. When examining the physical and the mental component summary, the difference was 0.99 and 2.72, respectively, the latter difference being statistically significant and with an associated effect size at 0.25.

Table 2. The 36-Item Short Form Health Survey scores in the internet and telephone groups after matching (N=245 each).

Scale or component summary and group	Score, median (IQR)	Score, mean (SD)	Score, mean (95% CI)	Score difference (telephone–internet)		
				P value	Mean difference	Effect size
Physical functioning				.02	4.57	0.18
Internet	85 (65-95)	76.08 (24.56)	76.08 (72.92-79.08)			
Telephone	90 (70-100)	80.65 (24.93)	80.65 (77.47-83.71)			
Role-physical				.002	9.39	0.22
Internet	50 (0-100)	51.53 (41.67)	51.53 (46.22-56.73)			
Telephone	100 (0-100)	60.92 (44.59)	60.92 (55.31-66.43)			
Bodily pain				.045	4.56	0.16
Internet	72 (41-100)	66.84 (26.12)	66.84 (63.55-70.11)			
Telephone	84 (41-100)	71.40 (32.23)	71.40 (67.42-75.40)			
General health				.99	0.04	0.00
Internet	57 (42-72)	55.10 (20.47)	55.10 (52.57-57.65)			
Telephone	57 (37-77)	55.15 (25.90)	55.15 (51.96-58.34)			
Vitality				.57	0.82	0.04
Internet	50 (35-65)	48.29 (20.16)	48.29 (45.78-50.80)			
Telephone	50 (35-65)	49.10 (21.30)	49.10 (46.41-51.78)			
Social functioning				<.001	7.96	0.29
Internet	75 (50-100)	71.17 (24.27)	71.17 (68.16-74.18)			
Telephone	100 (62.5-100)	79.13 (31.24)	79.13 (75.15-82.96)			
Role-emotional				.002	9.93	0.25
Internet	100 (33.33-100)	67.89 (39.04)	67.89 (63.13-72.65)			
Telephone	100 (66.66-100)	77.82 (39.84)	77.82 (72.79-82.59)			
Mental health				.002	5.01	0.26
Internet	64 (52-80)	63.56 (18.77)	63.56 (61.21-65.91)			
Telephone	72 (56-84)	68.57 (20.20)	68.57 (65.94-71.10)			
Physical component summary				.18	0.99	0.09
Internet	44.95 (37.27-53.30)	44.48 (10.04)	44.48 (43.20-45.75)			
Telephone	48.33 (37.81-54.43)	45.47 (11.05)	45.47 (44.09-46.82)			
Mental component summary				.02	2.72	0.25
Internet	47.49 (35.37-52.60)	44.68 (10.62)	44.68 (43.34-46.01)			
Telephone	50.86 (41.81-55.50)	47.40 (11.15)	47.40 (46.01-48.76)			

Discussion

The Opportunity to Investigate the Influence of the Mode of Administration of the Questionnaire on SF-36 Scores

To our knowledge, this study is the first reported to date to compare SF-36 questionnaire data collected either via telephone interviews or via self-completion on a dedicated internet website. More precisely, the availability of SF-36 data collected in the SENTIPAT trial provided a perfect opportunity to precisely investigate the influence of the mode of administration of the questionnaire on SF-36 scores. This investigation was the aim of the ancillary study of the SENTIPAT trial reported here and constitutes the major contribution of our report. This investigation has benefited from 3 main strengths. First, the study is based on a randomized trial with a substantial number of patients included in both arms. Second, the population under study had a large patient case mix variability because of the fact that patients originated from 5 very different hospital wards. The third strength of the study is the construction of a matched subsample of comparable responders in the 2 arms according to baseline variables to mitigate the impact of an unavoidable selection bias of responders as much as possible.

Principal Findings

Figure 2 shows that the matching procedure highly succeeded in composing a sample of similar match-paired patients; however, the very modest impact of this matching procedure on modifying the initial score differences between the scores in the internet and telephone groups (Figure 3) highly suggests that the score differences between internet and telephone groups are mainly attributable to the mode of administration, strictly speaking, with a minor impact of selection bias issues. In addition, and importantly, the scores in the telephone group were always higher than those in the internet group (Figure 3; Table 2), likely reflecting another type of bias associated with the telephone interview mode of administration: the interviewer effect.

For all but 2 out of 8 scales, the mean difference in scores between the groups was statistically significant and >4.5 points (Table 2), and several comments have to be made about this statement. It is worth recalling that misinterpretations of P values are very common [36,37]. A statistically significant score difference was not systematically considered as meaningful by the authors [38,39], and Ware et al [33] had initially proposed a 5-point difference between 2 SF-36 scores as a threshold value for a clinically and socially relevant difference. In our view, considering effect size is an appropriate approach for examining the relevance of score differences as such a perspective takes into account the variability of the measures and not only a rough mean difference threshold. Interestingly, as shown in Table 2, even if there were substantial mean score differences between the 2 different modes of administration in most of the scales, these differences were all related to a small effect size in 8 scales and in 2 component summaries of SF-36 according to the effect size index classification proposed by Cohen [34]. Cohen defined the small effect size as “noticeably smaller than medium which represents an effect visible to naked eye of a careful observer

but also not so small as to be trivial.” On the one hand, the effect size perspective considerably softens the observed differences between the internet and telephone groups and raises concerns about the relevance of considering a mean difference of 5 points as the main critical element of comparison between 2 scores. Moreover, our analyses also indicate that in studies involving a substantially variable population, only very large mean differences in scores would be considered meaningful when adopting the effect size perspective, highly limiting the usefulness of SF-36 in such studies. On the other hand, some mean differences in scores observed in our study and most likely attributable to the interviewer effect are not negligible. For example, in patients with chronic hepatitis C, Younossi et al [40] reported a mean value of the RP scale at 74.4 and 79.6 in patients with advanced and none to mild fibrosis, respectively ($P=.002$). Therefore, the differences for the RP scale, likely attributable to the SF-36 mode of administration observed in this study (51.5 and 60.9 in the internet and telephone groups, respectively; $P=.002$; Table 2), are at least comparable with those attributable to substantially different health states reported in other studies.

Limitations

The main limitation of the study concerns the selection bias related to responder status in both arms; however, such a bias is inherent to the 2 corresponding modes of administration, and this bias is likely different from one mode of administration to the other. In this study, selection biases were mitigated as much as possible by conducting a part of the analyses in a matched subpopulation of responders. A detailed analysis comparing the scores observed in the whole set of responders (before matching) and in a subpopulation enhancing the similarity of the compared individuals (after matching) constitutes an important strength of the study. Our results evidence an interviewer effect, which artificially increased SF-36 scores when the questionnaire was administered through a telephone interview. Therefore, the telephone interview as a mode of administration of SF-36 cumulates two types of bias: the unavoidable associated selection bias of responders and the interviewer effect, which is discussed in more detail in the following sections. In general, several methods can be used for mitigating the selection bias of responders as much as possible: one takes advantage of the distribution of baseline values observed in the responders and nonresponders to correct initial responder estimates to estimates more representative of the whole population under study [41]. In contrast, the interviewer effect raises many more concerns as the corresponding bias cannot be removed.

For the rest, some of the estimates reported here raise concerns in terms of generalizability and should only be viewed as minor side results that were required in the global process of the main goal of the study, which was to investigate the impact of the mode of administration of the SF-36 questionnaire on the collected scores. For example, the response rates reported here should not be considered emblematic of the corresponding modes of administration of the questionnaires. As detailed below in the *Response Rates According to the Mode of Administration* subsection, response rates reported in any study, including ours, are hardly generalizable as such rates likely depend on many characteristics of the survey design. Similarly, the reader should

keep in mind that the SF-36 scores collected here reflect the QoL of a particular population of patients admitted in 5 departments of a French university hospital, and these scores are not generalizable to other populations.

Comparison With Prior Works

Interviewer Effect

To our knowledge, this study is the only one to date that compared modes of administration of SF-36 on a matched sample of responders to mitigate—as much as possible—the inherent lack of initial comparability of responders according to the mode of administration of the questionnaire. Nevertheless, our results are in agreement with previous studies that reported higher SF-36 scores when administered by telephone than those issued from a mailed paper mode of administration [18,21,22,24-26]. Similarly, Lyons et al [42,43] reported higher scores issued from a face-to-face interview administration than those issued from a mailed paper self-completion of the SF-36 questionnaire. Altogether, our results and those of previous studies suggest that as compared with patients' self-completion, the introduction of an interviewer likely acts as a veil that somehow embellishes patients' QoL-reported perception. Internet self-completion avoids any potential bias of responses related to an interviewer effect [44], and patients are more likely to freely express their opinions [45] on websites covering anonymity than through telephone. Therefore, self-completion (internet or paper) should be preferred for collecting SF-36 data, as the involvement of a third party appears to artificially increase the scores. In any case, our study indicates that an accurate comparison of different scores requires at least avoiding modes of administration of SF-36 mixing self-completion and interview.

Response Rates According to the Mode of Administration

Despite the reminders sent to the patients, the internet group response rate (245/840, 29.2%) to the survey was dramatically lower than that of the telephone group (630/840, 75%). Blumenberg and Barros [46] explored the response rate differences between web and alternative data collection methods for public health research; considering the 9 papers comparing web self-completion with telephone and with a sample size >100, which were selected in their review, the median and range of the response rates reported for web and telephone were 23% and 2% to 68% and 40% and 8% to 71%, respectively. Similarly, a recent meta-analysis comparing response rates of web surveys with those obtained with other modes of administration [47] indicated that the results were stable when compared with a similar analysis conducted 10 years earlier: web surveys still yielded lower response rates than other modes, with a mean difference of 12% and large heterogeneity in the differences observed. No study compared telephone and internet administration modes for SF-36; however, the participation rates reported in studies that compared several modes of administration of SF-36 substantially varied from one study to another. For example, the response rate with the telephone was significantly higher than that with postal mail in the study by Wettergren et al [26] (77% and 63%, respectively; $P<.001$), as well as in the study by Perkins and Sanson-Fisher [24] (85%

and 68%, respectively; $P<.001$), whereas corresponding response rates were similar in the study by McHorney et al [23] (65.3% and 65.1%, respectively; $P=.68$) and in the study by Bursik et al [18] (71% and 68%, respectively; $P=.48$). In addition, the participation rate observed in our study in the internet group was close to that of Basnov et al [13], who reported a lower response rate in the internet group than that observed in the paper group (23% vs 76%, respectively).

In our view, the numeric value of the difference between the response rates observed in the 2 modes of administration of the present survey should be considered as a minor side result. Indeed, the heterogeneity of the comparisons reported in reviews [46,47] mostly reflects the fact that the differences between response rates collected via the internet versus other methods of administration reported in any survey are difficult to interpret and are not generalizable at all: the modes of administration include underlying elements of the whole survey process for which the impact on participation rate is hardly assessable or even describable, such as the internet website design in terms of its attractiveness or convenience or the detailed procedure for reaching participants by telephone. For example, the relatively high rate of participation in the telephone group observed in this study is likely related to the fact that the schedule of the telephone interview was arranged with each participant at the moment of his or her enrollment, and moreover, up to 3 calls were tried whenever the participant was not reached at the first phone call. In addition, many other features, such as the age distribution of the target population of the survey, might influence the observed response rates according to the mode or modes of administration of the questionnaire. In the end, internet use and use of telephones have evolved considerably since the completion of the SENTIPAT study. Such changes include not only technological aspects but also the growing importance of the abovementioned uses for many purposes throughout society. In particular, the COVID-19 crisis had a substantial impact on such matters [48,49]. Therefore, these changes are an additional element for limiting the interest in the intrinsic value of response rates, and it would be hazardous to consider that the participation rates in any web and telephone survey made before the COVID-19 period would be replicable whenever a similar survey would be conducted nowadays.

Conclusions

As compared with the mode of administration based on telephone interviews, the response rate of volunteer patients communicating their SF-36 data via the internet was much lower; however, our study indicates that a substantial proportion of hospitalized patients volunteered to actively document their health data via the internet. Most of all, the study indicates that the telephone interviewer might be viewed as an intermediate subjective pattern in the collection of patient data, resulting in a nonnegligible increase in SF-36 scores. Therefore, self-administration of SF-36 should be preferred, including via the internet, which is likely a low-cost method. Importantly, the results of this study also strongly advocate avoiding the conduction of surveys combining methods of SF-36 administration that mix self-reporting and interviews.

Acknowledgments

The Assistance Publique-Hôpitaux de Paris (Département de la Recherche Clinique et du Développement) was the trial sponsor. The SENTIPAT study was funded by grant AOM09213 K081204 from Programme Hospitalier de Recherche Clinique 2009 (Ministère de la Santé).

The sponsor and the funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

The authors would like to thank the SENTIPAT study group (scientific committee: Fabrice Carrat, Bérengère Couturier, Gilles Hejblum, Morgane Le Bail, Alain-Jacques Valleron, and the below-mentioned heads and physicians within the poles and units in which patients were recruited), personnel of the study group within the poles and units concerned (heads: Marc Beaussier, Jean-Paul Cabane, Olivier Chazouillères, Jacques Cosnes, Jean-Claude Dussaule, Pierre-Marie Girard, Emmanuel Tiret, and Dominique Pateron), additional corresponding physicians (Laurent Beaugerie, Laurence Fardet, Laurent Fonquernie, François Paye, and Laure Surgers), and nursing supervisors (Françoise Cuiller, Catherine Esnouf, Hélène Haure, Valérie Garnier, Josselin Mehal-Birba, Nelly Sallée, and Sylvie Wagener).

The authors are indebted to the excellent technical team of the study—the clinical research technicians (Élodie Belladame, Azéline Chevance, Magali Girard, and Laurence Nicole, who included patients, collected baseline data, and interviewed the patients followed-up by telephone) and the software staff (especially Pauline Raballand but also Frédéric Chau and Frédéric Fotré, who created and maintained the trial's dedicated website).

The authors would like to thank all the medical and nursing and administrative staff of the General and Digestive Surgery (including Ambulatory Surgery), Gastroenterology, Hepatology, Infectious Diseases, and Internal Medicine departments of Hôpital Saint-Antoine.

The authors would like to thank all patients who participated in the study.

Authors' Contributions

GH had full access to all the raw data in the study and can take responsibility for the integrity of the data and the accuracy of the data analysis. GH was involved in the study conception and design and data acquisition. AA and GH were involved in the analysis and writing of the first draft of the manuscript. AA, FC, and GH were involved in the interpretation of data. All authors approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Internal reliability of the 36-Item Short Form Health Survey in the internet and telephone groups.

[[PDF File \(Adobe PDF File\), 65 KB - jmir_v24i3e29009_app1.pdf](#)]

Multimedia Appendix 2

CONSORT-eHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 10460 KB - jmir_v24i3e29009_app2.pdf](#)]

References

1. Busija L, Pausenberger E, Haines TP, Haymes S, Buchbinder R, Osborne RH. Adult measures of general health and health-related quality of life: medical outcomes study short form 36-item (SF-36) and short form 12-item (SF-12) health surveys, Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), Medical Outcomes Study Short Form 6D (SF-6D), Health Utilities Index Mark 3 (HUI3), Quality of Well-Being Scale (QWB), and Assessment of Quality of Life (AQoL). *Arthritis Care Res (Hoboken)* 2011 Nov;63 Suppl 11:S383-S412 [[FREE Full text](#)] [doi: [10.1002/acr.20541](https://doi.org/10.1002/acr.20541)] [Medline: [22588759](https://pubmed.ncbi.nlm.nih.gov/22588759/)]
2. Contopoulos-Ioannidis DG, Karvouni A, Kouri I, Ioannidis JPA. Reporting and interpretation of SF-36 outcomes in randomised trials: systematic review. *BMJ* 2009;338:a3006 [[FREE Full text](#)] [Medline: [19139138](https://pubmed.ncbi.nlm.nih.gov/19139138/)]
3. Ware JE. SF-36 health survey update. *Spine (Phila Pa 1976)* 2000 Dec 15;25(24):3130-3139. [Medline: [11124729](https://pubmed.ncbi.nlm.nih.gov/11124729/)]
4. Leplège A, Ecosse E, Verdier A, Perneger TV. The French SF-36 Health Survey: translation, cultural adaptation and preliminary psychometric evaluation. *J Clin Epidemiol* 1998 Nov;51(11):1013-1023. [doi: [10.1016/s0895-4356\(98\)00093-6](https://doi.org/10.1016/s0895-4356(98)00093-6)] [Medline: [9817119](https://pubmed.ncbi.nlm.nih.gov/9817119/)]

5. Counsell SR, Callahan CM, Clark DO, Tu W, Buttar AB, Stump TE, et al. Geriatric care management for low-income seniors: a randomized controlled trial. *JAMA* 2007 Dec 12;298(22):2623-2633. [doi: [10.1001/jama.298.22.2623](https://doi.org/10.1001/jama.298.22.2623)] [Medline: [18073358](https://pubmed.ncbi.nlm.nih.gov/18073358/)]
6. Husky MM, Ferdous Farin F, Compagnone P, Fermanian C, Kovess-Masfety V. Chronic back pain and its association with quality of life in a large French population survey. *Health Qual Life Outcomes* 2018 Sep 26;16(1):195 [FREE Full text] [doi: [10.1186/s12955-018-1018-4](https://doi.org/10.1186/s12955-018-1018-4)] [Medline: [30257670](https://pubmed.ncbi.nlm.nih.gov/30257670/)]
7. Middleton S, McElduff P, Ward J, Grimshaw JM, Dale S, D'Este C, QASC Trialists Group. Implementation of evidence-based treatment protocols to manage fever, hyperglycaemia, and swallowing dysfunction in acute stroke (QASC): a cluster randomised controlled trial. *Lancet* 2011 Nov 12;378(9804):1699-1706. [doi: [10.1016/S0140-6736\(11\)61485-2](https://doi.org/10.1016/S0140-6736(11)61485-2)] [Medline: [21996470](https://pubmed.ncbi.nlm.nih.gov/21996470/)]
8. Needham DM, Dinglas VD, Bienvenu OJ, Colantuoni E, Wozniak AW, Rice TW, NIH NHLBI ARDS Network. One year outcomes in patients with acute lung injury randomised to initial trophic or full enteral feeding: prospective follow-up of EDEN randomised trial. *BMJ* 2013 Mar 19;346:f1532 [FREE Full text] [doi: [10.1136/bmj.f1532](https://doi.org/10.1136/bmj.f1532)] [Medline: [23512759](https://pubmed.ncbi.nlm.nih.gov/23512759/)]
9. Bell DS, Kahn CE. Health status assessment via the World Wide Web. *Proc AMIA Annu Fall Symp* 1996:338-342 [FREE Full text] [Medline: [8947684](https://pubmed.ncbi.nlm.nih.gov/8947684/)]
10. Cunha-Miranda L, Santos H, Miguel C, Silva C, Barcelos F, Borges J, et al. Validation of Portuguese-translated computer touch-screen questionnaires in patients with rheumatoid arthritis and spondyloarthritis, compared with paper formats. *Rheumatol Int* 2015 Dec;35(12):2029-2035. [doi: [10.1007/s00296-015-3347-5](https://doi.org/10.1007/s00296-015-3347-5)] [Medline: [26346588](https://pubmed.ncbi.nlm.nih.gov/26346588/)]
11. Hidalgo-Mazzei D, Reinares M, Mateu A, Nikolova VL, Bonnín CD, Samalin L, et al. OpenSIMPLE: a real-world implementation feasibility study of a smartphone-based psychoeducation programme for bipolar disorder. *J Affect Disord* 2018 Dec 01;241:436-445. [doi: [10.1016/j.jad.2018.08.048](https://doi.org/10.1016/j.jad.2018.08.048)] [Medline: [30145515](https://pubmed.ncbi.nlm.nih.gov/30145515/)]
12. Kao H, Wu W, Liang T, Lee K, Hou M, Shi H. Cloud-based service information system for evaluating quality of life after breast cancer surgery. *PLoS One* 2015;10(9):e0139252 [FREE Full text] [doi: [10.1371/journal.pone.0139252](https://doi.org/10.1371/journal.pone.0139252)] [Medline: [26422018](https://pubmed.ncbi.nlm.nih.gov/26422018/)]
13. Basnov M, Kongsved SM, Bech P, Hjollund NH. Reliability of short form-36 in an internet- and a pen-and-paper version. *Inform Health Soc Care* 2009 Jan;34(1):53-58. [doi: [10.1080/17538150902779527](https://doi.org/10.1080/17538150902779527)] [Medline: [19306199](https://pubmed.ncbi.nlm.nih.gov/19306199/)]
14. Broering JM, Paciorek A, Carroll PR, Wilson LS, Litwin MS, Miaskowski C. Measurement equivalence using a mixed-mode approach to administer health-related quality of life instruments. *Qual Life Res* 2014 Mar;23(2):495-508. [doi: [10.1007/s11136-013-0493-7](https://doi.org/10.1007/s11136-013-0493-7)] [Medline: [23943258](https://pubmed.ncbi.nlm.nih.gov/23943258/)]
15. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of internet versus paper-and-pencil versions. *J Med Internet Res* 2007;9(3):e25 [FREE Full text] [doi: [10.2196/jmir.9.3.e25](https://doi.org/10.2196/jmir.9.3.e25)] [Medline: [17942387](https://pubmed.ncbi.nlm.nih.gov/17942387/)]
16. MacKenzie H, Thavaneswaran A, Chandran V, Gladman DD. Patient-reported outcome in psoriatic arthritis: a comparison of web-based versus paper-completed questionnaires. *J Rheumatol* 2011 Dec;38(12):2619-2624. [doi: [10.3899/jrheum.110165](https://doi.org/10.3899/jrheum.110165)] [Medline: [22045844](https://pubmed.ncbi.nlm.nih.gov/22045844/)]
17. Naus MJ, Philipp LM, Samsi M. From paper to pixels: a comparison of paper and computer formats in psychological assessment. *Comput Hum Behav* 2009 Jan;25(1):1-7. [doi: [10.1016/j.chb.2008.05.012](https://doi.org/10.1016/j.chb.2008.05.012)]
18. Buskirk TD, Stein KD. Telephone vs. mail survey gives different SF-36 quality-of-life scores among cancer survivors. *J Clin Epidemiol* 2008 Oct;61(10):1049-1055. [doi: [10.1016/j.jclinepi.2007.11.012](https://doi.org/10.1016/j.jclinepi.2007.11.012)] [Medline: [18538997](https://pubmed.ncbi.nlm.nih.gov/18538997/)]
19. Cerrada CJ, Weinberg J, Sherman KJ, Saper RB. Inter-method reliability of paper surveys and computer assisted telephone interviews in a randomized controlled trial of yoga for low back pain. *BMC Res Notes* 2014 Apr 09;7:227 [FREE Full text] [doi: [10.1186/1756-0500-7-227](https://doi.org/10.1186/1756-0500-7-227)] [Medline: [24716775](https://pubmed.ncbi.nlm.nih.gov/24716775/)]
20. García M, Rohlfes I, Vila J, Sala J, Pena A, Masiá R, REGICOR Investigators. Comparison between telephone and self-administration of Short Form Health Survey Questionnaire (SF-36). *Gac Sanit* 2005;19(6):433-439 [FREE Full text] [doi: [10.1016/s0213-9111\(05\)71393-5](https://doi.org/10.1016/s0213-9111(05)71393-5)] [Medline: [16483520](https://pubmed.ncbi.nlm.nih.gov/16483520/)]
21. Hays RD, Kim S, Spritzer KL, Kaplan RM, Tally S, Feeny D, et al. Effects of mode and order of administration on generic health-related quality of life scores. *Value Health* 2009 Sep;12(6):1035-1039 [FREE Full text] [doi: [10.1111/j.1524-4733.2009.00566.x](https://doi.org/10.1111/j.1524-4733.2009.00566.x)] [Medline: [19473334](https://pubmed.ncbi.nlm.nih.gov/19473334/)]
22. Jörggård A, Wettergen L, von Essen L. Measuring health-related quality of life in adolescents and young adults: Swedish normative data for the SF-36 and the HADS, and the influence of age, gender, and method of administration. *Health Qual Life Outcomes* 2006 Dec 01;4:91 [FREE Full text] [doi: [10.1186/1477-7525-4-91](https://doi.org/10.1186/1477-7525-4-91)] [Medline: [17140436](https://pubmed.ncbi.nlm.nih.gov/17140436/)]
23. McHorney CA, Kosinski M, Ware JE. Comparisons of the costs and quality of norms for the SF-36 health survey collected by mail versus telephone interview: results from a national survey. *Med Care* 1994 Jun;32(6):551-567. [Medline: [8189774](https://pubmed.ncbi.nlm.nih.gov/8189774/)]
24. Perkins JJ, Sanson-Fisher RW. An examination of self- and telephone-administered modes of administration for the Australian SF-36. *J Clin Epidemiol* 1998 Nov;51(11):969-973. [Medline: [9817114](https://pubmed.ncbi.nlm.nih.gov/9817114/)]
25. Powers JR, Mishra G, Young AF. Differences in mail and telephone responses to self-rated health: use of multiple imputation in correcting for response bias. *Aust N Z J Public Health* 2005 Apr;29(2):149-154. [doi: [10.1111/j.1467-842x.2005.tb00065.x](https://doi.org/10.1111/j.1467-842x.2005.tb00065.x)] [Medline: [15915619](https://pubmed.ncbi.nlm.nih.gov/15915619/)]

26. Wettergren L, Mattsson E, von Essen L. Mode of administration only has a small effect on data quality and self-reported health status and emotional distress among Swedish adolescents and young adults. *J Clin Nurs* 2011 Jun;20(11-12):1568-1577. [doi: [10.1111/j.1365-2702.2010.03481.x](https://doi.org/10.1111/j.1365-2702.2010.03481.x)] [Medline: [21323779](https://pubmed.ncbi.nlm.nih.gov/21323779/)]
27. Sentinel patients: value of an information system collecting patient's own report on his healthcare management (SENTIPAT). NIH U.S. National Library of Medicines. URL: <https://clinicaltrials.gov/ct2/show/NCT01769261> [accessed 2020-12-17]
28. Couturier B, Carrat F, Hejblum G. Comparing patients' opinions on the hospital discharge process collected with a self-reported questionnaire completed via the internet or through a telephone survey: an ancillary study of the SENTIPAT randomized controlled trial. *J Med Internet Res* 2015 Jun 24;17(6):e158 [FREE Full text] [doi: [10.2196/jmir.4379](https://doi.org/10.2196/jmir.4379)] [Medline: [26109261](https://pubmed.ncbi.nlm.nih.gov/26109261/)]
29. Feldman SF, Lapidus N, Cosnes J, Turet E, Fonquernie L, Cabane J, et al. Comparing inpatient satisfaction collected via a web-based questionnaire self-completion and through a telephone interview: an ancillary study of the SENTIPAT randomized controlled trial. *J Med Internet Res* 2017 Aug 23;19(8):e293 [FREE Full text] [doi: [10.2196/jmir.7061](https://doi.org/10.2196/jmir.7061)] [Medline: [28835354](https://pubmed.ncbi.nlm.nih.gov/28835354/)]
30. Leplège A, Ecosse E, Pouchot J, Coste J, Perneger T. Le questionnaire MOS SF-36: manuel de l'utilisateur et guide d'interprétation des scores. Paris, France: Éditions ESTEM; 2001.
31. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol* 2017 Feb;82:128-136 [FREE Full text] [doi: [10.1016/j.jclinepi.2016.11.016](https://doi.org/10.1016/j.jclinepi.2016.11.016)] [Medline: [27986610](https://pubmed.ncbi.nlm.nih.gov/27986610/)]
32. Rouquette A, Blanchin M, Sébille V, Guillemin F, Côté SM, Falissard B, et al. The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline score. *J Clin Epidemiol* 2014 Apr;67(4):433-440. [doi: [10.1016/j.jclinepi.2013.10.009](https://doi.org/10.1016/j.jclinepi.2013.10.009)] [Medline: [24447591](https://pubmed.ncbi.nlm.nih.gov/24447591/)]
33. Ware J, Snow K, Kosinski M, Gandek B. SF-36 Health Survey: Manual and Interpretation Guide. Lincoln, RI: Quality Metric Incorporated; 1993.
34. Cohen J. A power primer. *Psychol Bull* 1992 Jul;112(1):155-159. [Medline: [19565683](https://pubmed.ncbi.nlm.nih.gov/19565683/)]
35. Ho DE, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference. *J Stat Soft* 2011;42(8):1-28. [doi: [10.18637/jss.v042.i08](https://doi.org/10.18637/jss.v042.i08)]
36. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016 May 21;31(4):337-350. [doi: [10.1007/s10654-016-0149-3](https://doi.org/10.1007/s10654-016-0149-3)]
37. Nuzzo R. Scientific method: statistical errors. *Nature* 2014 Feb 13;506(7487):150-152. [doi: [10.1038/506150a](https://doi.org/10.1038/506150a)] [Medline: [24522584](https://pubmed.ncbi.nlm.nih.gov/24522584/)]
38. Hays J, Ockene JK, Brunner RL, Kotchen JM, Manson JE, Patterson RE, Women's Health Initiative Investigators. Effects of estrogen plus progestin on health-related quality of life. *N Engl J Med* 2003 May 08;348(19):1839-1854. [doi: [10.1056/NEJMoa030311](https://doi.org/10.1056/NEJMoa030311)] [Medline: [12642637](https://pubmed.ncbi.nlm.nih.gov/12642637/)]
39. Yost KJ, Haan MN, Levine RA, Gold EB. Comparing SF-36 scores across three groups of women with different health profiles. *Qual Life Res* 2005 Jun;14(5):1251-1261. [doi: [10.1007/s11136-004-6673-8](https://doi.org/10.1007/s11136-004-6673-8)] [Medline: [16047501](https://pubmed.ncbi.nlm.nih.gov/16047501/)]
40. Younossi ZM, Stepanova M, Afdhal N, Kowdley KV, Zeuzem S, Henry L, et al. Improvement of health-related quality of life and work productivity in chronic hepatitis C patients with early and advanced fibrosis treated with ledipasvir and sofosbuvir. *J Hepatol* 2015 Aug;63(2):337-345. [doi: [10.1016/j.jhep.2015.03.014](https://doi.org/10.1016/j.jhep.2015.03.014)] [Medline: [25795586](https://pubmed.ncbi.nlm.nih.gov/25795586/)]
41. Lumley T. *Complex Surveys: A Guide to Analysis Using R*. Oxford, UK: Wiley-Blackwell; 2010:1-296.
42. Lyons RA, Fielder H, Littlepage BN. Measuring health status with the SF-36: the need for regional norms. *J Public Health Med* 1995 Mar;17(1):46-50. [Medline: [7786567](https://pubmed.ncbi.nlm.nih.gov/7786567/)]
43. Lyons RA, Wareham K, Lucas M, Price D, Williams J, Hutchings HA. SF-36 scores vary by method of administration: implications for study design. *J Public Health Med* 1999 Mar;21(1):41-45. [doi: [10.1093/pubmed/21.1.41](https://doi.org/10.1093/pubmed/21.1.41)] [Medline: [10321858](https://pubmed.ncbi.nlm.nih.gov/10321858/)]
44. Dijkstra W. How interviewer variance can bias the results of research on interviewer effects. *Qual Quant* 1983 Jun;17(3):179-187. [doi: [10.1007/BF00167582](https://doi.org/10.1007/BF00167582)]
45. Kreuter F, Presser S, Tourangeau R. Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. *Public Opin Quart* 2009 Jan 26;72(5):847-865. [doi: [10.1093/poq/nfn063](https://doi.org/10.1093/poq/nfn063)]
46. Blumenberg C, Barros AJ. Response rate differences between web and alternative data collection methods for public health research: a systematic review of the literature. *Int J Public Health* 2018 Jul;63(6):765-773. [doi: [10.1007/s00038-018-1108-4](https://doi.org/10.1007/s00038-018-1108-4)] [Medline: [29691594](https://pubmed.ncbi.nlm.nih.gov/29691594/)]
47. Daikeler J, Bosnjak M, Manfreda K. Web versus other survey modes: an updated and extended meta-analysis comparing response rates. *J Surv Stat Methodol* 2020;8(3):513-539. [doi: [10.1093/jssam/smz008](https://doi.org/10.1093/jssam/smz008)]
48. Colley RC, Bushnik T, Langlois K. Exercise and screen time during the COVID-19 pandemic. *Health Rep* 2020 Jul 15;31(6):3-11 [FREE Full text] [doi: [10.25318/82-003-x202000600001-eng](https://doi.org/10.25318/82-003-x202000600001-eng)] [Medline: [32672923](https://pubmed.ncbi.nlm.nih.gov/32672923/)]
49. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, RADAR-CNS Consortium. Using smartphones and wearable devices to monitor behavioral changes during COVID-19. *J Med Internet Res* 2020 Sep 25;22(9):e19992 [FREE Full text] [doi: [10.2196/19992](https://doi.org/10.2196/19992)] [Medline: [32877352](https://pubmed.ncbi.nlm.nih.gov/32877352/)]

Abbreviations**QoL:** quality of life**RP:** role-physical**SF-36:** 36-Item Short Form Health Survey

Edited by A Mavragani; submitted 22.03.21; peer-reviewed by S Bidmon, S Pranic, N Dreyer; comments to author 21.07.21; revised version received 21.10.21; accepted 21.12.21; published 10.03.22.

Please cite as:

Açma A, Carrat F, Hejblum G

Comparing SF-36 Scores Collected Through Web-Based Questionnaire Self-completions and Telephone Interviews: An Ancillary Study of the SENTIPAT Multicenter Randomized Controlled Trial

J Med Internet Res 2022;24(3):e29009

URL: <https://www.jmir.org/2022/3/e29009>

doi: [10.2196/29009](https://doi.org/10.2196/29009)

PMID: [35266869](https://pubmed.ncbi.nlm.nih.gov/35266869/)

©Ayşe Açma, Fabrice Carrat, Gilles Hejblum. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 10.03.2022. 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.

Original Paper

Development of a Prognostic App (iCanPredict) to Predict Survival for Chinese Women With Breast Cancer: Retrospective Study

Zhuo Ma^{1*}, BSc; Sijia Huang^{1*}, BSc; Xiaoqing Wu^{2*}, BSc; Yinying Huang^{3*}, MSc; Sally Wai-Chi Chan⁴, PhD; Yilan Lin², MSc; Xujuan Zheng⁵, PhD; Jiemin Zhu¹, PhD

¹Department of Nursing, School of Medicine, Xiamen University, Xiamen, China

²Department of Chronic Non-infectious Diseases and Endemic Diseases Control, Xiamen Center for Disease Control and Prevention, Xiamen, China

³Department of Nursing, Women and Children's Hospital, School of Medicine, Xiamen University, Xiamen, China

⁴President Office, Tung Wah College, Hong Kong, China

⁵School of Nursing, Health Science Centre, Shenzhen University, Shenzhen, China

*these authors contributed equally

Corresponding Author:

Jiemin Zhu, PhD

Department of Nursing

School of Medicine

Xiamen University

Room 222, Alice Lee Building, Xiang'an Campus, Xiamen University

Xiang'an South Road

Xiamen, 361005

China

Phone: 86 15960212649

Email: jieminzhu@xmu.edu.cn

Abstract

Background: Accurate prediction of survival is crucial for both physicians and women with breast cancer to enable clinical decision making on appropriate treatments. The currently available survival prediction tools were developed based on demographic and clinical data obtained from specific populations and may underestimate or overestimate the survival of women with breast cancer in China.

Objective: This study aims to develop and validate a prognostic app to predict the overall survival of women with breast cancer in China.

Methods: Nine-year (January 2009-December 2017) clinical data of women with breast cancer who received surgery and adjuvant therapy from 2 hospitals in Xiamen were collected and matched against the death data from the Xiamen Center of Disease Control and Prevention. All samples were randomly divided (7:3 ratio) into a training set for model construction and a test set for model external validation. Multivariable Cox regression analysis was used to construct a survival prediction model. The model performance was evaluated by receiver operating characteristic (ROC) curve and Brier score. Finally, by running the survival prediction model in the app background thread, the prognostic app, called iCanPredict, was developed for women with breast cancer in China.

Results: A total of 1592 samples were included for data analysis. The training set comprised 1114 individuals and the test set comprised 478 individuals. Age at diagnosis, clinical stage, molecular classification, operative type, axillary lymph node dissection, chemotherapy, and endocrine therapy were incorporated into the model, where age at diagnosis (hazard ratio [HR] 1.031, 95% CI 1.011-1.051; $P=.002$), clinical stage (HR 3.044, 95% CI 2.347-3.928; $P<.001$), and endocrine therapy (HR 0.592, 95% CI 0.384-0.914; $P=.02$) significantly influenced the survival of women with breast cancer. The operative type ($P=.81$) and the other 4 variables (molecular classification [$P=.91$], breast reconstruction [$P=.36$], axillary lymph node dissection [$P=.32$], and chemotherapy [$P=.84$]) were not significant. The ROC curve of the training set showed that the model exhibited good discrimination for predicting 1- (area under the curve [AUC] 0.802, 95% CI 0.713-0.892), 5- (AUC 0.813, 95% CI 0.760-0.865), and 10-year (AUC 0.740, 95% CI 0.672-0.808) overall survival. The Brier scores at 1, 5, and 10 years after diagnosis were 0.005, 0.055, and 0.103 in the training set, respectively, and were less than 0.25, indicating good predictive ability. The test set externally validated

model discrimination and calibration. In the iCanPredict app, when physicians or women input women's clinical information and their choice of surgery and adjuvant therapy, the corresponding 10-year survival prediction will be presented.

Conclusions: This survival prediction model provided good model discrimination and calibration. iCanPredict is the first tool of its kind in China to provide survival predictions to women with breast cancer. iCanPredict will increase women's awareness of the similar survival rate of different surgeries and the importance of adherence to endocrine therapy, ultimately helping women to make informed decisions regarding treatment for breast cancer.

(*J Med Internet Res* 2022;24(3):e35768) doi:[10.2196/35768](https://doi.org/10.2196/35768)

KEYWORDS

app; breast cancer; survival prediction model; iCanPredict

Introduction

Accurate prediction of survival will help both physicians and women with breast cancer to determine the type of surgery and the adjuvant therapy that are beneficial for these women [1]. Physicians usually make survival predictions and adjuvant therapy formulations according to the clinical characteristics of women with breast cancer, such as age, tumor size, clinical stage, hormone receptor status, and molecular characteristics, in addition to their own clinical experience [1-3]. The predicted adjuvant benefit is achieved by calculating the percentage improvement in the predicted survival rate after adjuvant treatment [1].

Currently, some survival prediction tools are widely used, such as PREDICT [1,4] and Adjuvant! Online [5-9]. PREDICT was developed based on clinical data of women with breast cancer collected from East Anglia (UK) between 1999 and 2003 [1,4], and Adjuvant! Online was developed based on data at the Netherlands Cancer Institute from 1987 to 1998 [5-9]. Studies have shown that such prediction tools underestimated or overestimated the survival rate of women with breast cancer, especially in Asia [10-12]. Differences in access to health care and the quality of existing health care among countries are possible reasons [13].

In China, the practice of surgery for breast cancer markedly differs from that in Western countries. A national survey reported that breast-conserving surgery accounted for 22% of surgeries in China [14], in contrast to 64.5% in the United States [15], despite the similar survival rates of breast-conserving surgery and mastectomy for women with early stage breast cancer [16,17]. Furthermore, some recent studies have shown that breast-conserving surgery has higher breast cancer-specific and overall survival rates than mastectomy [18]. The insufficient and unbalanced medical resources, information disparity between physicians and women with breast cancer, distrust of breast-conserving surgery by some physicians and women, and fear of recurrence are important reasons for the low rate of breast-conserving surgery in China [14]. In addition, for women with breast cancer in China who had received adjuvant endocrine therapy, the nonpersistence rates ranged from 5.9% to 22.6% in years 1-5, and the rate of low compliance ranged from 2.1% to 11.2% [19]. The low compliance with endocrine therapy may be partially explained by inadequate information and social support [19-23]. Therefore, demonstrating the similar survival rate of different surgeries and the efficacy of adherence to

endocrine therapy may help physicians and women with breast cancer make better clinical decisions.

In 2020, the popularity of mobile phones in China exceeded 1 per person [24], and 920 million people could surf the internet using mobile phones [25]. Further, WeChat, a popular mobile phone platform, has over 1 billion monthly active users in China [26]. The Mini app, one of the functionalities of WeChat, provides an easily accessible platform to present the Chinese survival prediction model for women with breast cancer.

To the best of our knowledge, no prognostic app has been designed and applied to women with breast cancer in China. Such survival prediction tools would be beneficial to help physicians and women with breast cancer obtain and understand the survival rate of different surgeries and adjuvant therapy, thereby enabling better clinical decision making. The aim of this study is to develop and validate a prognostic app to predict the overall survival of women with breast cancer in China.

Methods

Study Design

A retrospective design was applied. All samples were randomly divided (7:3 ratio) into a training set for model construction and internal validation, and a test set for model external validation [27].

Study Setting and Population

Clinical data collection was conducted at 2 hospitals in Xiamen, a tertiary hospital and a hospital for women and children health. Data on death were collected from the Xiamen Center of Disease Control and Prevention (CDC), which retains all death records for Xiamen residents.

The inclusion criteria were women with breast cancer (1) who received treatment including surgery and adjuvant therapy in these 2 hospitals from January 1, 2009, to December 31, 2017; and (2) who were Xiamen residents. Women who lacked surgical and treatment information were excluded from the analysis.

Sample Size Calculation

The pmsampsize package for Stata and R written by Riley et al [28] was used to calculate the sample size. The 10-year mortality rate in women with breast cancer was 0.214, and the median follow-up was 11.7 years [6]. To target a margin of error in the estimate of death risk of ≤ 0.05 , a total of 680 samples were

needed with Cox-Snell R^2 statistic of 0.1 [28], assumed outcome event rate of 0.214, follow-up of 11.7 years [6], timepoint of interest for 10 years' prediction, and 8 parameters (variables).

Procedure

The researchers collected clinical data of all patients with breast cancer who received treatment at 2 participating hospitals from January 2009 to December 2017. Further, the population information on death for Xiamen residents from January 2009 to December 2020 was obtained from the Xiamen Center CDC. The citizen ID number was used as the basis for matching. If the match was successful, the death dates of those women were recorded. If not successful, women with breast cancer were assumed to be alive. Finally, a data set was developed for model construction and validation.

Ethical Consideration

Ethical approvals were obtained from Research Ethics Committees at School of Medicine, Xiamen University (number XDYX2019008), First Hospital Affiliated to Xiamen University (number XMDY-2020-019), and Women and Children's Hospital affiliated to Xiamen University (number KY-2019-058). To protect the privacy of women with breast cancer, data collection, entry, and processing were carried out by 2 professional researchers. During the use of iCanPredict, the users' input information can only be accessed by authorized researchers, who needed to log-in to the iCanPredict website background thread with the protected account number and passport. Accordingly, unauthorized personnel was prohibited from viewing and using the clinical data and users' input information.

Data Collected

Outcomes data were obtained from the medical records (January 2009-December 2017) of the 2 participating hospitals including age at diagnosis, clinical stage (stage 0-IV), molecular classification (luminal A, luminal B, Her-2 overexpression, basal-like), operative type (breast conservative surgery, mastectomy), breast reconstruction, axillary lymph node dissection, chemotherapy, and endocrine therapy. These 2 participating hospitals used the TNM clinical stages from the American Joint Committee on Cancer, which are determined by tumor size (T), lymph node metastasis (N), and distant metastasis (M) [29].

Xiamen residents' death information from January 1, 2009, to December 15, 2020, was collected from the Xiamen Center CDC. Women's citizen ID number was retrieved from the hospital to link women's medical records to Xiamen residents' death information at the Xiamen Center CDC.

Outcomes

The 5-year (or 10-year) mortality was defined as the number of deaths divided by the total number of breast cancer cases from the diagnosis to 5-year (or 10-year) follow-up. The outcome event was the death of women with breast cancer from January 1, 2009, to December 15, 2020. The survival of women with breast cancer at the end of follow-up was considered as a truncated event [30].

Data Analysis

Data analysis was performed using SPSS version 25.0 (IBM) [31] and R version 4.0.4 (R Foundation). Demographic data were summarized using descriptive statistics. Because all variables satisfied the proportional hazards (PH) assumption, multivariate Cox regression analysis was used to develop a PH regression model for time-to-event data (eg, happening of death) and produce an equation to predict an individual's outcome risk on the condition of her values of multiple predictors [28].

Model Construction

The training set was used for model construction. On the basis of the hazard ratio (HR) of each prognosis factor estimated by multivariable Cox regression, the survival prediction model was expressed as

$$S(t) = S_0(t) \exp(-\beta X)$$

where $S(t)$ indicates the predicted survival rate of women with breast cancer after treatment, $S_0(t)$ represents the baseline survival rate of women with breast cancer, β is the coefficient for each prognostic factor in the multivariable Cox regression, X is the value of the prognosis factor, and \bar{X} is the mean of the prognosis factor.

Validation

The training set was analyzed for model internal validation and the test set was analyzed for model external validation. Model discrimination was evaluated using the receiver operating characteristic (ROC) curve. In general, the larger the area under the curve (AUC), the higher the model discrimination. Further, an AUC greater than 0.7 indicates that the model has a certain distinctive ability. Model calibration was based on the Brier score; the lower the Brier score, the higher the calibration degree. When the Brier score was less than 0.25, the model could be predicted.

Design of a Prognostic App to Predict Survival

The research team comprised researchers, statisticians, breast cancer specialists, and app specialists at Xiamen University, the participating hospitals, and Tung Wah College of Hong Kong. The Xiamen Quanwu Information Service Company (Xiamen, People's Republic of China) undertook the technical development and maintenance of the iCanPredict app.

iCanPredict is mainly divided into a website background thread (back end) and the user side (front end). The website background thread was developed to approve the users' applications for registration, run the survival prediction model on the basis of the users' input information, and track the app usage frequency and duration. Only researchers are able to access the website background thread with protected account number and passport. The users' side was developed to interact with users and present the 10-year survival rate for women with breast cancer.

According to the Technology Acceptance Model proposed by Davis [32], perceived ease of use and perceived usefulness are 2 decisive variables of the user's reception of the information system. Perceived ease of use refers to the degree of difficulty a user perceives when using an information system [33].

Perceived usefulness refers to the fact that users can obtain certain values when using an information system [33]. Perceived ease of use affects perceived usefulness and further affects users' actual use behavior by affecting their attitudes and behaviors [33]. To promote ease of use, the iCanPredict app was incorporated into the functionality of WeChat, the most popular mobile phone platform in China. Prior to implementing the iCanPredict app, the perceived ease of use and perceived usefulness were tested by women with breast cancer and surgeons.

Results

Participants' Demographic and Clinic Characteristics

The clinical data of 1686 participants were collected, of which 97 lacked treatment information, leaving data on 1592 participants for model construction and validation. The training set comprised 1114 individuals and the test set comprised 478

individuals. By December 15, 2020, 147/1592 participants died (9.23%) and 1445/1592 (90.77%) were still alive. In this study, the mean follow-up was 6.38 years, ranging from 0.32 to 12.35 years. The mean age at diagnosis was 49.92 (SD 11.59) years, and most participants were diagnosed with stage II breast cancer (733/1592, 46.04%). Luminal B-type (1059/1592, 66.52%) was the most common molecular classification. Only 95/1592 participants (5.97%) underwent breast-conserving surgery, while 1497/1592 (94.03%) underwent mastectomy. A total of 453/1497 participants underwent breast reconstruction, accounting for 30.26% of the participants who underwent mastectomy and 28.45% (453/1592) of all participants. Most participants underwent axillary lymph node dissection (1143/1592, 71.80%), received adjuvant chemotherapy (1305/1592, 81.97%), and received adjuvant endocrine therapy (1007/1592, 63.25%). The demographic and clinical characteristics of the participants in the training and test sets are presented in [Table 1](#).

Table 1. Demographic and clinical characteristics of participants in the training set and the test set.^a

Characteristics	Total (n=1592)	Training set (n=1114)	Test set (n=478)
Deaths, n (%)	147 (9.23)	103 (9.25)	44 (9.21)
Patients alive, n (%)	1445 (90.77)	1011 (90.75)	434 (90.79)
Follow-up (years), mean (SD)	6.38 (2.68)	6.40 (2.68)	6.32 (2.68)
Age at diagnosis (years), mean (SD)	49.92 (11.59)	49.68 (11.48)	50.50 (11.82)
Clinical stage, n (%)			
0	26 (1.63)	21 (1.89)	5 (1.05)
I	524 (32.91)	363 (32.59)	161 (33.68)
II	733 (46.04)	508 (45.60)	225 (47.07)
III	286 (17.96)	204 (18.31)	82 (17.15)
IV	23 (1.44)	18 (1.62)	5 (1.05)
Molecular classification, n (%)			
Luminal A	215 (13.51)	160 (14.36)	55 (11.51)
Luminal B	1059 (66.52)	723 (64.90)	336 (70.29)
HER-2 (+)	232 (14.57)	168 (15.08)	64 (13.39)
Basal like	86 (5.40)	63 (5.66)	23 (4.81)
Operative type, n (%)			
Breast-conserving surgery	95 (5.97)	68 (6.10)	27 (5.65)
Mastectomy	1497 (94.03)	1046 (93.90)	451 (94.35)
Breast reconstruction, n (%)			
Yes	453 (28.45)	321 (28.82)	132 (27.62)
No	1139 (71.55)	793 (71.18)	346 (72.38)
Axillary lymph node dissection, n (%)			
Yes	1143 (71.80)	810 (72.71)	333 (69.67)
No	449 (28.20)	304 (27.29)	145 (30.33)
Chemotherapy, n (%)			
Yes	1305 (81.97)	919 (82.50)	386 (80.75)
No	287 (18.03)	195 (17.50)	92 (19.25)
Endocrine therapy, n (%)			
Yes	1007 (63.25)	707 (63.46)	300 (62.76)
No	585 (36.75)	407 (36.54)	178 (37.24)

^aPercentage=number/group total number.

Initial Model Fit

The training set was used to fit the initial model. The graph based on the standardized Schoenfeld residual method and the $-\ln(-\ln[\text{survival}])$ test indicated that all 8 variables met the PH assumption. A survival prediction model was constructed for women with breast cancer (Table 2). Age at diagnosis ($P=.002$), clinical stage ($P<.001$), and endocrine therapy ($P=.02$) were identified as significant prognostic factors. The risk of death increased by 0.031 times per year with an increase in age at

diagnosis for women with breast cancer (HR 1.031, 95% CI 1.011-1.051). When the clinical stage increased, the risk of death in women increased by 2.044 times per stage (HR 3.044, 95% CI 2.347-3.928). Women with breast cancer receiving endocrine therapy had a 0.407 (1–0.592) times lower risk of death than those that did not receive endocrine therapy (HR 0.592, 95% CI 0.384-0.914). The other 5 variables, including molecular classification ($P=.91$), operative type ($P=.81$), breast reconstruction ($P=.36$), axillary lymph node dissection ($P=.32$), and chemotherapy ($P=.84$), were not significant.

Table 2. Hazard ratios and model coefficients for prognostic factors included in the models (n=1114).

Prognostic factors	β	Standard error	Hazard ratio (95% CI)	P value
Age at diagnosis	0.031	0.01	1.031 (1.011-1.051)	.002
Clinical stage	1.113	0.133	3.044 (2.347-3.928)	<.001
Molecular classification	0.016	0.145	1.017 (0.765-1.351)	.91
Operative type	0.127	0.52	1.136 (0.410-3.145)	.81
Breast reconstruction	-0.318	0.343	0.728 (0.372-1.426)	.36
Axillary lymph node dissection	0.42	0.422	1.521 (0.665-3.478)	.32
Chemotherapy	-0.067	0.34	0.935 (0.480-1.821)	.84
Endocrine therapy	-0.524	0.222	0.592 (0.384-0.914)	.02

Validation Results

All statistically significant and nonsignificant factors are included in the prediction model because all factors are clinically significant. Further, the prediction model including nonsignificant factors presented model discrimination and calibration similar to those presented by the prediction model excluding nonsignificant factors. The AUC and Brier score of the training and test sets are presented in [Tables 3](#) and [4](#), respectively. For internal validation with the training set, the ROC curve of the training set indicated that the model exhibited good discrimination for predicting 1- (AUC 0.802, 95% CI

0.713-0.892), 5- (AUC 0.813, 95% CI 0.760-0.865), and 10-year (AUC 0.740, 95% CI 0.672-0.808) overall survival ([Figure 1A,C,E](#)). The Brier scores at 1, 5, and 10 years after diagnosis were 0.005, 0.055, and 0.103, respectively. For external validation with the test set, except that the AUC at 10 years after diagnosis was 0.685, the AUC values 1 and 5 years after diagnosis were greater than 0.7, indicating that the model has a certain distinctive ability ([Figure 1B,D,F](#)). The Brier scores 1, 5, and 10 years after diagnosis were less than 0.25, indicating that the model calibration was also good. Therefore, the prediction model could be used to predict survival.

Table 3. Internal validation and external validation: model discrimination (AUC^a) at 1, 5, and 10 years after diagnosis.

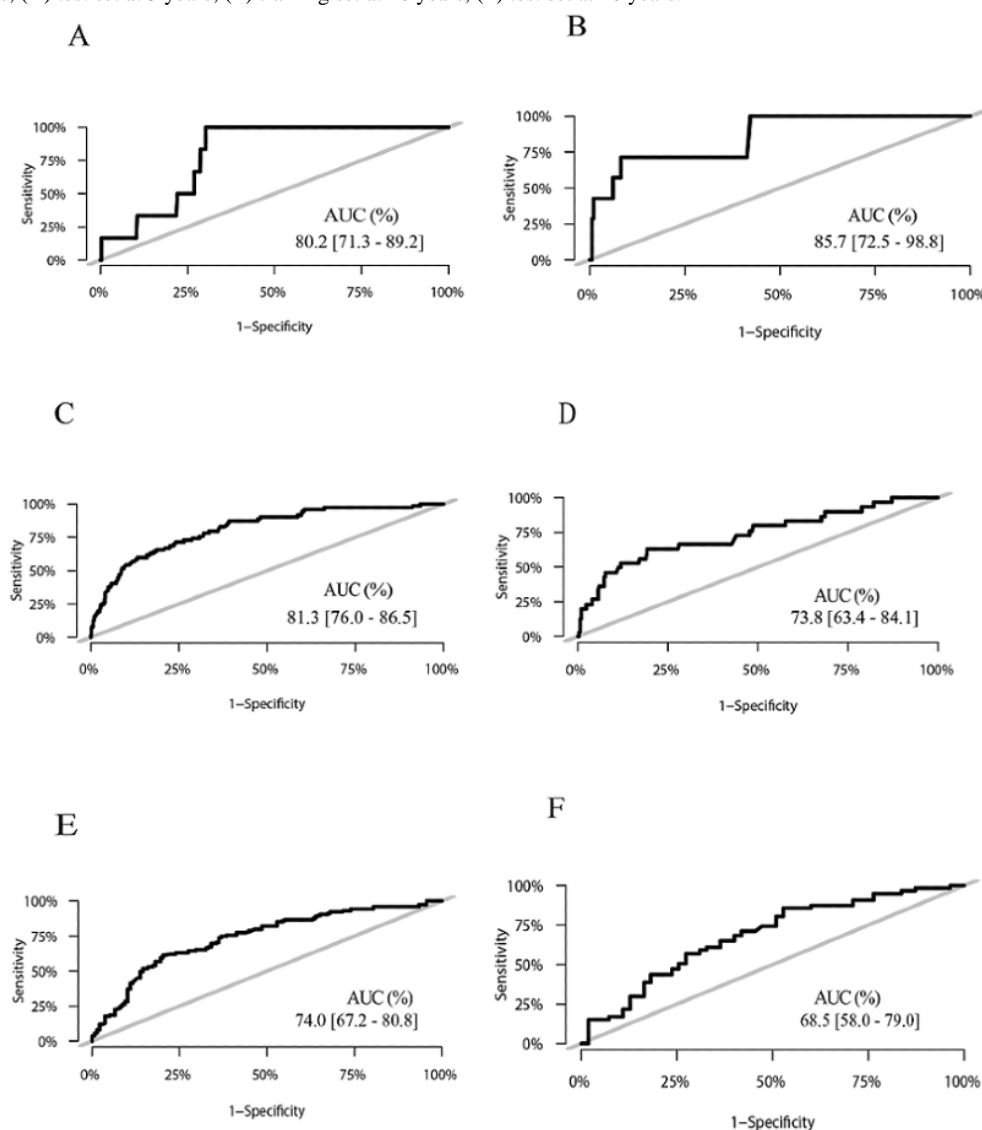
Year	Internal validation		External validation	
	AUC	95% CI	AUC	95% CI
1	0.802	0.713-0.892	0.857	0.725-0.988
5	0.813	0.760-0.865	0.738	0.634-0.841
10	0.740	0.672-0.808	0.685	0.580-0.790

^aAUC: area under the curve.

Table 4. Internal validation and external validation: calibration (Brier) at 1, 5, and 10 years after diagnosis.

Year	Internal validation		External validation	
	Brier score	95% CI	Brier score	95% CI
1	0.005	0.001-0.010	0.014	0.004-0.025
5	0.055	0.043-0.067	0.057	0.038-0.075
10	0.103	0.083-0.124	0.120	0.084-0.156

Figure 1. Receiver operator characteristic (ROC) curves for breast cancer overall survival rates. (A) Training set at 1 year; (B) test set at 1 year; (C) training set at 5 years; (D) test set at 5 years; (E) training set at 10 years; (F) test set at 10 years.



Development of Preliminary Survival Prediction for Breast Cancer Survivors

The development of iCanPredict was based on a uniAPP open-source free framework with separation technology for the back end and the front end. The back end was developed using the Java language 8 environment, IntelliJ IDEA for code development, and Maven for service dependency management. The front end utilized an Ant-design-VUE component for page interaction construction under the VUE framework. Each function in the iCanPredict app was tested multiple times to ensure usability and stability.

A QR code was generated for the users to scan with their own WeChat account. After downloading the iCanPredict app into their WeChat platform, the users need apply for registration. The researcher will approve the users' application from the

website background thread. Each user's mobile phone number is set up as the unique username and an automated passport is sent to the user's mobile phone (changeable later). The users do not need to pay for the access to the iCanPredict app.

iCanPredict includes 2 function modules: survival prediction and a personal home page for women with breast cancer. The purpose and data sources are displayed on the home page with a statement as follows: "iCanPredict aims to be used for scientific research and provide the survival probability of different treatment options. Neither the model author nor the relevant hospitals can guarantee the calculation accuracy for any specific patient. Any decision related to cancer treatment should consult your physician". Women need to click the button "I agree" under the statement before using iCanPredict (Figure 2A).

Figure 2. Screenshots of the iCanPredict app.



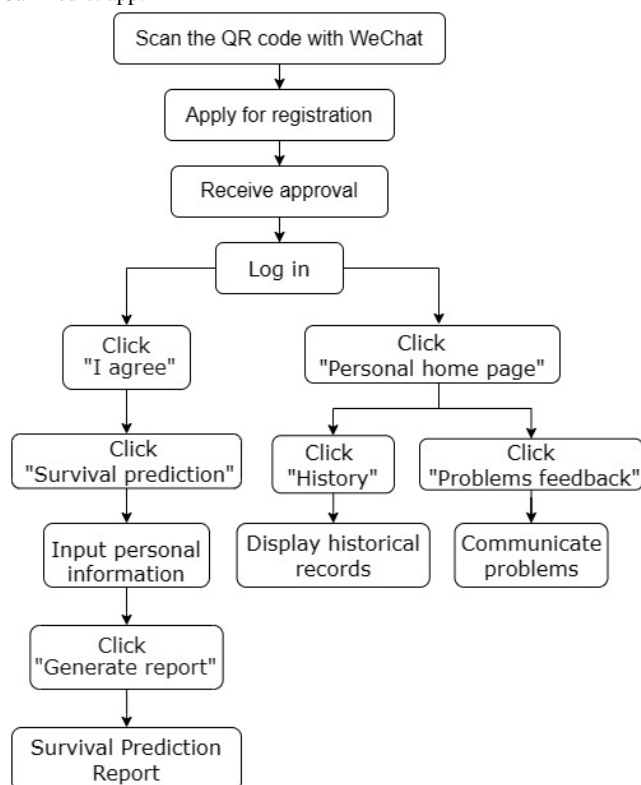
In the survival prediction model module, there are interactive page and report page. In the interactive page, users need to input personal information, including the year of birth, the year of diagnosis, breast cancer clinical stage, and molecular classification. In addition, users need to choose surgery type (breast-conserving surgery or mastectomy), whether they will undergo reconstruction if they choose mastectomy, whether they will undergo axillary lymph node dissection, and whether they will receive chemotherapy or endocrine therapy (Figure 2B). In addition, to make it easier and faster for users to learn how to input information, there are tutorials and concept explanations beside each concept in the input information interface. Tutorials and concept explanations can resolve doubts in the use process. After the completion of information input, users need to click the button “generate report” at the bottom of the interactive page.

In the report page, iCanPredict will present the 10-year survival rates after surgery only, surgery plus chemotherapy, and surgery plus chemotherapy plus endocrine therapy based on the user’s input information. The survival rates are presented as tables and diagrams for women with breast cancer. The tables present 5- and 10-year survival rates, while the diagrams describe the continuous survival rates within 10 years after primary treatment. The predicted benefits of adjuvant chemotherapy and endocrine therapy are represented by the percentage increase in the predicted survival rate with differently colored curves in the diagram for women with breast cancer who received adjuvant therapy, compared with those who did not receive adjuvant therapy [1] (Figure 2C).

Users can make different choices on different surgeries and adjuvant therapies, and the corresponding survival prediction data will be saved on their personal home page. The personal home page module includes personal centers, problems

feedback, and history. Personal center refers to the account-related information (nickname, avatar, basic information, etc.), which is set up by the users themselves. Problems feedback is an interactive platform for communication between users and the iCanPredict developer group. Users can provide feedback on the problems encountered in the use of iCanPredict or provide suggestions for app improvement. History summarizes user input information and survival prediction data are expressed in tables and diagrams and arranged in chronological order for users to compare the survival rates of different surgeries and adjuvant treatments. The flowchart of the use of iCanPredict is presented in Figure 3.

A total of 6 women with breast cancer and 2 surgeons were invited to assess the perceived ease of use and perceived usefulness. Of these, 2 women were aged over 60 years and were reluctant to use the iCanPredict app. One woman got confused about her own value of molecular classification and needed to refer to the laboratory test report to input information. With help from the researchers to download the iCanPredict app and explain the laboratory test, all 6 women showed great interest in the prediction model. They appraised the iCanPredict app’s convenience, ease of use, and ability to provide valuable information they wanted. The 2 surgeons indicated that it was easy to use the iCanPredict app. The surgeons expressed surprise that mastectomy and breast conservative surgery resulted in similar survival rates. They also suggested to add more information to predict survivals following other treatment options, such as immunotherapy or gene therapies. Such therapies have been emphasized to treat women with breast cancer in the recent years, but were not commonly applied before. Both women with breast cancer and surgeons indicated that it was feasible and acceptable to use the iCanPredict app in the clinics.

Figure 3. Flowchart of the use of the iCanPredict app.

Discussion

Principal Findings

In this study, we constructed a well-calibrated prediction model on the basis of 9 years of clinical and death information for women with breast cancer in Xiamen. The survival prediction model was run in the app background thread, and the iCanPredict prognostic app was designed and developed for women with breast cancer in China.

iCanPredict was developed on the popular mobile phone platform in China (WeChat), which is easily accessible and user-friendly for both physicians and women with breast cancer [34]. iCanPredict provides individual survival predictions and calculates the benefits of the corresponding treatments. The tables display the percentage of predicted survival rates for physicians and women to compare the predicted results among different surgeries and adjuvant therapies. The use of colored diagrams to display the improvement in the predicted survival rate will facilitate sophisticated and difficult discussions on corresponding adjuvant therapies for patients [1].

Comparison With Prior Work

In our prediction model, the type of surgery was not an independent prognostic factor for breast cancer overall survival, which is consistent with some previous studies [16,17]. When physicians discuss the choices of breast surgeries with women, they may input women's clinical data and attempt different types of surgery in the iCanPredict app, which presents the survival prediction rate for different surgeries in an easy-to-understand manner. In combination with value preferences, women may thus make better informed decisions on surgeries [35,36].

In our prediction model, endocrine therapy was a significant prognostic factor for overall survival, which was also supported by the literature [37]. iCanPredict will present the benefit of endocrine therapy based on the percentage of improvement in the predicted survival rate if physicians or women choose endocrine therapy. iCanPredict will provide a quick, visual presentation, such as tables and diagrams, to inform women about the beneficial effects of endocrine therapy.

For clinical implementation, it is recommended that women with breast cancer use the iCanPredict together with surgeons when making decisions on treatments. Further, cautions should be exercised when using this iCanPredict app to guide the decision making, considering new therapies that are constantly emerging in recent years such as immunotherapy [38-41] or gene-based therapies [42,43], which may affect the survival rate of women with breast cancer.

Strengths and Limitations

iCanPredict is the first tool of its kind in China to provide survival predictions to women scheduled to undergo breast surgeries and corresponding adjuvant therapies. iCanPredict was proved to have good model discrimination and calibration. This study had several limitations. First, the sample size was relatively small to build a prediction model, as there were only limited number of eligible cases in the 2 participating hospitals, which may limit the generalization of the iCanPredict app. Second, death data from the Xiamen Center CDC did not include specific causes of death; therefore, this study did not have specific breast cancer survival rates, and only contained overall survival rates. Therefore, the iCanPredict app may overestimate the mortality of breast cancer. Third, due to the incomplete clinical records of adjuvant endocrine therapy and adjuvant chemotherapy treatment, we can only predict the benefits of

adjuvant therapy by determining whether adjuvant therapy was administered (yes vs no). In future studies, the adherence to adjuvant endocrine therapy should be emphasized to help clinicians and patients make accurate and individual medical decisions.

Future Directions

Future studies need to combine the results of this study with the financial burden of different treatments [44] and patients' own preferences [35,36] to provide better clinical decision support. Studies with large sample size and more comprehensive clinical data, in combination with the medical costs related to

medical treatments and patients' value preference, are warranted to be conducted in different countries and cultures.

Conclusions

The prediction model had good model discrimination and model calibration to predict the overall survival of women with breast cancer in China. iCanPredict will increase women's awareness of the similar survival rate of different surgeries and the importance of adherence to endocrine therapy, ultimately helping them make informed decisions about treatment for breast cancer. In the long run, better choice of surgery and increased adherence to prescribed endocrine therapy may be attained.

Acknowledgments

The authors thank physicians and nurses, and other individuals that assisted with data collection and processing. This research was funded by the National Natural Science Foundation of China (grant number 71974162). The funding body had no role in the study design, data collection and interpretation, and manuscript preparation.

Authors' Contributions

JZ and XZ designed the study. JZ and SWCC obtained the funding for this study. SH, XW, YL, YH, and XZ collected the data. ZM analyzed and interpreted the data. ZM, JZ, and XZ interpreted the results and wrote the manuscript. All authors read the manuscript.

Conflicts of Interest

None declared.

References

1. Wishart GC, Azzato EM, Greenberg DC, Rashbass J, Kearins O, Lawrence G, et al. PREDICT: a new UK prognostic model that predicts survival following surgery for invasive breast cancer. *Breast Cancer Res* 2010;12(1):R1. [doi: [10.1186/bcr2464](https://doi.org/10.1186/bcr2464)] [Medline: [20053270](https://pubmed.ncbi.nlm.nih.gov/20053270/)]
2. Wang X, Feng Z, Huang Y, Li H, Cui P, Wang D, et al. A Nomogram To Predict The Overall Survival Of Breast Cancer Patients And Guide The Postoperative Adjuvant Chemotherapy In China. *Cancer Manag Res* 2019;11:10029-10039 [FREE Full text] [doi: [10.2147/CMAR.S215000](https://doi.org/10.2147/CMAR.S215000)] [Medline: [31819635](https://pubmed.ncbi.nlm.nih.gov/31819635/)]
3. Michaelson JS, Chen LL, Bush D, Fong A, Smith B, Younger J. Improved web-based calculators for predicting breast carcinoma outcomes. *Breast Cancer Res Treat* 2011 Aug;128(3):827-835. [doi: [10.1007/s10549-011-1366-9](https://doi.org/10.1007/s10549-011-1366-9)] [Medline: [21327471](https://pubmed.ncbi.nlm.nih.gov/21327471/)]
4. Candido Dos Reis FJ, Wishart GC, Dicks EM, Greenberg D, Rashbass J, Schmidt MK, et al. An updated PREDICT breast cancer prognostication and treatment benefit prediction model with independent validation. *Breast Cancer Res* 2017 May 22;19(1):58 [FREE Full text] [doi: [10.1186/s13058-017-0852-3](https://doi.org/10.1186/s13058-017-0852-3)] [Medline: [28532503](https://pubmed.ncbi.nlm.nih.gov/28532503/)]
5. Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, et al. Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer. *J Clin Oncol* 2001 Feb 15;19(4):980-991. [doi: [10.1200/JCO.2001.19.4.980](https://doi.org/10.1200/JCO.2001.19.4.980)] [Medline: [11181660](https://pubmed.ncbi.nlm.nih.gov/11181660/)]
6. Mook S, Schmidt MK, Rutgers EJ, van de Velde AO, Visser O, Rutgers SM, et al. Calibration and discriminatory accuracy of prognosis calculation for breast cancer with the online Adjuvant! program: a hospital-based retrospective cohort study. *Lancet Oncol* 2009 Nov;10(11):1070-1076. [doi: [10.1016/S1470-2045\(09\)70254-2](https://doi.org/10.1016/S1470-2045(09)70254-2)] [Medline: [19801202](https://pubmed.ncbi.nlm.nih.gov/19801202/)]
7. Campbell HE, Taylor MA, Harris AL, Gray AM. An investigation into the performance of the Adjuvant! Online prognostic programme in early breast cancer for a cohort of patients in the United Kingdom. *Br J Cancer* 2009 Oct 06;101(7):1074-1084 [FREE Full text] [doi: [10.1038/sj.bjc.6605283](https://doi.org/10.1038/sj.bjc.6605283)] [Medline: [19724274](https://pubmed.ncbi.nlm.nih.gov/19724274/)]
8. Olivotto IA, Bajdik CD, Ravdin PM, Speers CH, Coldman AJ, Norris BD, et al. Population-based validation of the prognostic model ADJUVANT! for early breast cancer. *J Clin Oncol* 2005 Apr 20;23(12):2716-2725. [doi: [10.1200/JCO.2005.06.178](https://doi.org/10.1200/JCO.2005.06.178)] [Medline: [15837986](https://pubmed.ncbi.nlm.nih.gov/15837986/)]
9. Wishart GC, Bajdik CD, Azzato EM, Dicks E, Greenberg DC, Rashbass J, et al. A population-based validation of the prognostic model PREDICT for early breast cancer. *Eur J Surg Oncol* 2011 May;37(5):411-417. [doi: [10.1016/j.ejso.2011.02.001](https://doi.org/10.1016/j.ejso.2011.02.001)] [Medline: [21371853](https://pubmed.ncbi.nlm.nih.gov/21371853/)]

10. Yao-Lung K, Dar-Ren C, Tsai-Wang C. Accuracy validation of adjuvant! online in Taiwanese breast cancer patients--a 10-year analysis. *BMC Med Inform Decis Mak* 2012 Sep 17;12:108 [FREE Full text] [doi: [10.1186/1472-6947-12-108](https://doi.org/10.1186/1472-6947-12-108)] [Medline: [22985190](https://pubmed.ncbi.nlm.nih.gov/22985190/)]
11. Bhoo-Pathy N, Yip C, Hartman M, Saxena N, Taib NA, Ho G, et al. Adjuvant! Online is overoptimistic in predicting survival of Asian breast cancer patients. *Eur J Cancer* 2012 May;48(7):982-989 [FREE Full text] [doi: [10.1016/j.ejca.2012.01.034](https://doi.org/10.1016/j.ejca.2012.01.034)] [Medline: [22366561](https://pubmed.ncbi.nlm.nih.gov/22366561/)]
12. de Glas NA, van de Water W, Engelhardt EG, Bastiaannet E, de Craen AJM, Kroep JR, et al. Validity of Adjuvant! Online program in older patients with breast cancer: a population-based study. *Lancet Oncol* 2014 Jun;15(7):722-729. [doi: [10.1016/S1470-2045\(14\)70200-1](https://doi.org/10.1016/S1470-2045(14)70200-1)] [Medline: [24836274](https://pubmed.ncbi.nlm.nih.gov/24836274/)]
13. Mazo C, Kearns C, Mooney C, Gallagher WM. Clinical Decision Support Systems in Breast Cancer: A Systematic Review. *Cancers (Basel)* 2020 Feb 06;12(2):1-15 [FREE Full text] [doi: [10.3390/cancers12020369](https://doi.org/10.3390/cancers12020369)] [Medline: [32041094](https://pubmed.ncbi.nlm.nih.gov/32041094/)]
14. Yang B, Ren G, Song E, Pan D, Zhang J, Wang Y, et al. Current Status and Factors Influencing Surgical Options for Breast Cancer in China: A Nationwide Cross-Sectional Survey of 110 Hospitals. *Oncologist* 2020 Oct;25(10):e1473-e1480 [FREE Full text] [doi: [10.1634/theoncologist.2020-0001](https://doi.org/10.1634/theoncologist.2020-0001)] [Medline: [32333626](https://pubmed.ncbi.nlm.nih.gov/32333626/)]
15. Kummerow KL, Du L, Penson DF, Shyr Y, Hooks MA. Nationwide trends in mastectomy for early-stage breast cancer. *JAMA Surg* 2015 Jan;150(1):9-16. [doi: [10.1001/jamasurg.2014.2895](https://doi.org/10.1001/jamasurg.2014.2895)] [Medline: [25408966](https://pubmed.ncbi.nlm.nih.gov/25408966/)]
16. Sinnadurai S, Kwong A, Hartman M, Tan EY, Bhoo-Pathy NT, Dahlui M, et al. Breast-conserving surgery mastectomy in young women with breast cancer in Asian settings. *BJS Open* 2019 Feb;3(1):48-55 [FREE Full text] [doi: [10.1002/bjs5.50111](https://doi.org/10.1002/bjs5.50111)] [Medline: [30734015](https://pubmed.ncbi.nlm.nih.gov/30734015/)]
17. van Maaren MC, de Munck L, de Bock GH, Jobsen JJ, van Dalen T, Linn SC, et al. 10 year survival after breast-conserving surgery plus radiotherapy compared with mastectomy in early breast cancer in the Netherlands: a population-based study. *Lancet Oncol* 2016 Aug;17(8):1158-1170. [doi: [10.1016/S1470-2045\(16\)30067-5](https://doi.org/10.1016/S1470-2045(16)30067-5)] [Medline: [27344114](https://pubmed.ncbi.nlm.nih.gov/27344114/)]
18. Legendijk M, van Maaren MC, Saadatmand S, Strobbe LJA, Poortmans PMP, Koppert LB, et al. Breast conserving therapy and mastectomy revisited: Breast cancer-specific survival and the influence of prognostic factors in 129,692 patients. *Int J Cancer* 2018 Jan 01;142(1):165-175 [FREE Full text] [doi: [10.1002/ijc.31034](https://doi.org/10.1002/ijc.31034)] [Medline: [28884470](https://pubmed.ncbi.nlm.nih.gov/28884470/)]
19. Xu H, Jin F, Zhang X, Wang D, Yu S, Wang A. Adherence status to Adjuvant Endocrine Therapy in Chinese Women with Early Breast Cancer and its influencing factors: A cross-sectional survey. *Cancer Med* 2020 Jun;9(11):3703-3713 [FREE Full text] [doi: [10.1002/cam4.3017](https://doi.org/10.1002/cam4.3017)] [Medline: [32237070](https://pubmed.ncbi.nlm.nih.gov/32237070/)]
20. Xu H, Zhang X, Wang D, Xu L, Wang A. Factors influencing medication-taking behaviour with adjuvant endocrine therapy in women with breast cancer: A qualitative systematic review. *J Adv Nurs* 2020 Feb;76(2):445-458. [doi: [10.1111/jan.14253](https://doi.org/10.1111/jan.14253)] [Medline: [31657028](https://pubmed.ncbi.nlm.nih.gov/31657028/)]
21. Simon R, Latreille J, Matte C, Desjardins P, Bergeron E. Adherence to adjuvant endocrine therapy in estrogen receptor-positive breast cancer patients with regular follow-up. *Can J Surg* 2014 Feb;57(1):26-32 [FREE Full text] [doi: [10.1503/cjs.006211](https://doi.org/10.1503/cjs.006211)] [Medline: [24461223](https://pubmed.ncbi.nlm.nih.gov/24461223/)]
22. Heisig SR, Shedden-Mora MC, von Blanckenburg P, Schuricht F, Rief W, Albert U, et al. Informing women with breast cancer about endocrine therapy: effects on knowledge and adherence. *Psychooncology* 2015 Feb;24(2):130-137. [doi: [10.1002/pon.3611](https://doi.org/10.1002/pon.3611)] [Medline: [24953538](https://pubmed.ncbi.nlm.nih.gov/24953538/)]
23. Friese CR, Pini TM, Li Y, Abrahamse PH, Graff JJ, Hamilton AS, et al. Adjuvant endocrine therapy initiation and persistence in a diverse sample of patients with breast cancer. *Breast Cancer Res Treat* 2013 Apr;138(3):931-939 [FREE Full text] [doi: [10.1007/s10549-013-2499-9](https://doi.org/10.1007/s10549-013-2499-9)] [Medline: [23542957](https://pubmed.ncbi.nlm.nih.gov/23542957/)]
24. Number of mobile cell phone subscriptions in China from December 2019 to December 2020 (in millions). Statista. URL: <https://www.statista.com/statistics/278204/china-mobile-users-by-month/> [accessed 2021-09-19]
25. Statista. Number of mobile internet users in China from 2015 to 2019 with a forecast until 2023 (in millions). Statista. URL: <https://www.statista.com/statistics/558731/number-of-mobile-internet-user-in-china/> [accessed 2021-09-19]
26. 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](https://doi.org/10.7717/peerj.10550)] [Medline: [33362979](https://pubmed.ncbi.nlm.nih.gov/33362979/)]
27. Zhu J, Hao J, Ma Q, Shi T, Wang S, Yan J, et al. A Novel Prognostic Model and Practical Nomogram for Predicting the Outcomes of Colorectal Cancer: Based on Tumor Biomarkers and Log Odds of Positive Lymph Node Scheme. *Front Oncol* 2021;11:661040 [FREE Full text] [doi: [10.3389/fonc.2021.661040](https://doi.org/10.3389/fonc.2021.661040)] [Medline: [33937076](https://pubmed.ncbi.nlm.nih.gov/33937076/)]
28. Riley RD, Ensor J, Snell KIE, Harrell FE, Martin GP, Reitsma JB, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ* 2020 Mar 18;368:m441. [doi: [10.1136/bmj.m441](https://doi.org/10.1136/bmj.m441)] [Medline: [32188600](https://pubmed.ncbi.nlm.nih.gov/32188600/)]
29. Giuliano AE, Connolly JL, Edge SB, Mittendorf EA, Rugo HS, Solin LJ, et al. Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* 2017 Jul 08;67(4):290-303 [FREE Full text] [doi: [10.3322/caac.21393](https://doi.org/10.3322/caac.21393)] [Medline: [28294295](https://pubmed.ncbi.nlm.nih.gov/28294295/)]
30. Clark TG, Bradburn MJ, Love SB, Altman DG. Survival analysis part I: basic concepts and first analyses. *Br J Cancer* 2003 Jul 21;89(2):232-238 [FREE Full text] [doi: [10.1038/sj.bjc.6601118](https://doi.org/10.1038/sj.bjc.6601118)] [Medline: [12865907](https://pubmed.ncbi.nlm.nih.gov/12865907/)]
31. IBM Corporation. IBM SPSS Statistics for Windows (25th edition). 2017. URL: <https://www.ibm.com/in-en/analytics/spss-statistics-software> [accessed 2022-03-04]

32. Davis F. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly* 1989 Sep 28;13(3):319-6333. [doi: [10.2307/249008](https://doi.org/10.2307/249008)] [Medline: [32716422](https://pubmed.ncbi.nlm.nih.gov/32716422/)]
33. 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](https://doi.org/10.1016/j.jbi.2009.07.002)] [Medline: [19615467](https://pubmed.ncbi.nlm.nih.gov/19615467/)]
34. Li Y, Ye S, Zhou Y, Mao F, Guo H, Lin Y, et al. Web-Based Medical Information Searching by Chinese Patients With Breast Cancer and its Influence on Survival: Observational Study. *J Med Internet Res* 2020 Apr 17;22(4):e16768 [FREE Full text] [doi: [10.2196/16768](https://doi.org/10.2196/16768)] [Medline: [32301740](https://pubmed.ncbi.nlm.nih.gov/32301740/)]
35. Hawley ST, Newman L, Griggs JJ, Kosir MA, Katz SJ. Evaluating a Decision Aid for Improving Decision Making in Patients with Early-stage Breast Cancer. *Patient* 2016 Apr;9(2):161-169 [FREE Full text] [doi: [10.1007/s40271-015-0135-y](https://doi.org/10.1007/s40271-015-0135-y)] [Medline: [26178202](https://pubmed.ncbi.nlm.nih.gov/26178202/)]
36. Minami CA, King TA, Mittendorf EA. Patient preferences for locoregional therapy in early-stage breast cancer. *Breast Cancer Res Treat* 2020 Sep;183(2):291-309. [doi: [10.1007/s10549-020-05737-9](https://doi.org/10.1007/s10549-020-05737-9)] [Medline: [32691377](https://pubmed.ncbi.nlm.nih.gov/32691377/)]
37. Davies C, Pan H, Godwin J, Gray R, Arriagada R, Raina V, Adjuvant Tamoxifen: Longer Against Shorter (ATLAS) Collaborative Group. Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet* 2013 Mar 09;381(9869):805-816 [FREE Full text] [doi: [10.1016/S0140-6736\(12\)61963-1](https://doi.org/10.1016/S0140-6736(12)61963-1)] [Medline: [23219286](https://pubmed.ncbi.nlm.nih.gov/23219286/)]
38. Basu A, Ramamoorthi G, Jia Y, Faughn J, Wiener D, Awshah S, et al. Immunotherapy in breast cancer: Current status and future directions. *Adv Cancer Res* 2019;143:295-349. [doi: [10.1016/bs.acr.2019.03.006](https://doi.org/10.1016/bs.acr.2019.03.006)] [Medline: [31202361](https://pubmed.ncbi.nlm.nih.gov/31202361/)]
39. Marra A, Viale G, Curigliano G. Recent advances in triple negative breast cancer: the immunotherapy era. *BMC Med* 2019 May 09;17(1):90 [FREE Full text] [doi: [10.1186/s12916-019-1326-5](https://doi.org/10.1186/s12916-019-1326-5)] [Medline: [31068190](https://pubmed.ncbi.nlm.nih.gov/31068190/)]
40. Emens LA. Immunotherapy in Triple-Negative Breast Cancer. *Cancer J* 2021;27(1):59-66. [doi: [10.1097/PPO.0000000000000497](https://doi.org/10.1097/PPO.0000000000000497)] [Medline: [33475294](https://pubmed.ncbi.nlm.nih.gov/33475294/)]
41. Zhu Y, Zhu X, Tang C, Guan X, Zhang W. Progress and challenges of immunotherapy in triple-negative breast cancer. *Biochim Biophys Acta Rev Cancer* 2021 Dec;1876(2):188593. [doi: [10.1016/j.bbcan.2021.188593](https://doi.org/10.1016/j.bbcan.2021.188593)] [Medline: [34280474](https://pubmed.ncbi.nlm.nih.gov/34280474/)]
42. Jakub JW, Peled AW, Gray RJ, Greenup RA, Kiluk JV, Sacchini V, et al. Oncologic Safety of Prophylactic Nipple-Sparing Mastectomy in a Population With BRCA Mutations: A Multi-institutional Study. *JAMA Surg* 2018 Feb 01;153(2):123-129 [FREE Full text] [doi: [10.1001/jamasurg.2017.3422](https://doi.org/10.1001/jamasurg.2017.3422)] [Medline: [28903167](https://pubmed.ncbi.nlm.nih.gov/28903167/)]
43. Diéras V, Han HS, Kaufman B, Wildiers H, Friedlander M, Ayoub J, et al. Veliparib with carboplatin and paclitaxel in BRCA-mutated advanced breast cancer (BROCADE3): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol* 2020 Oct;21(10):1269-1282. [doi: [10.1016/S1470-2045\(20\)30447-2](https://doi.org/10.1016/S1470-2045(20)30447-2)] [Medline: [32861273](https://pubmed.ncbi.nlm.nih.gov/32861273/)]
44. Fan L, Strasser-Weippl K, Li J, St LJ, Finkelstein DM, Yu K, et al. Breast cancer in China. *Lancet Oncol* 2014 Jun;15(7):e279-e289. [doi: [10.1016/S1470-2045\(13\)70567-9](https://doi.org/10.1016/S1470-2045(13)70567-9)] [Medline: [24872111](https://pubmed.ncbi.nlm.nih.gov/24872111/)]

Abbreviations

AUC: area under the curve

HR: hazard ratio

PH: proportional hazards

ROC: receiver operating characteristic

Xiamen Center CDC: Xiamen Center of Disease Control and Prevention

Edited by T Leung; submitted 21.12.21; peer-reviewed by H He, J Zhang; comments to author 11.01.22; revised version received 28.01.22; accepted 21.02.22; published 09.03.22.

Please cite as:

Ma Z, Huang S, Wu X, Huang Y, Chan SWC, Lin Y, Zheng X, Zhu J

Development of a Prognostic App (iCanPredict) to Predict Survival for Chinese Women With Breast Cancer: Retrospective Study
J Med Internet Res 2022;24(3):e35768

URL: <https://www.jmir.org/2022/3/e35768>

doi: [10.2196/35768](https://doi.org/10.2196/35768)

PMID: [35262503](https://pubmed.ncbi.nlm.nih.gov/35262503/)

©Zhuo Ma, Sijia Huang, Xiaoqing Wu, Yinying Huang, Sally Wai-Chi Chan, Yilan Lin, Xujuan Zheng, Jiemin Zhu. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 09.03.2022. 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.

Original Paper

Machine Learning Prediction Models for Gestational Diabetes Mellitus: Meta-analysis

Zheqing Zhang¹, BEng; Luqian Yang¹, MBBS; Wentao Han¹, MBBS; Yaoyu Wu¹, MBBS; Linhui Zhang¹, MBBS; Chun Gao¹, MBBS; Kui Jiang¹, MSc; Yun Liu^{2,3}, PhD; Huiqun Wu¹, PhD

¹Department of Medical Informatics, Medical School of Nantong University, Nantong, China

²Department of Information, The First Affiliated Hospital, Nanjing Medical University, Nanjing, China

³Department of Medical Informatics, School of Biomedical Engineering and Informatics, Nanjing Medical University, Nanjing, China

Corresponding Author:

Huiqun Wu, PhD

Department of Medical Informatics

Medical School of Nantong University

Chongchuan District

19 Qixiu Road

Nantong, 226001

China

Phone: 86 513 85051891

Fax: 86 513 85051820

Email: wuhuiqun@ntu.edu.cn

Abstract

Background: Gestational diabetes mellitus (GDM) is a common endocrine metabolic disease, involving a carbohydrate intolerance of variable severity during pregnancy. The incidence of GDM-related complications and adverse pregnancy outcomes has declined, in part, due to early screening. Machine learning (ML) models are increasingly used to identify risk factors and enable the early prediction of GDM.

Objective: The aim of this study was to perform a meta-analysis and comparison of published prognostic models for predicting the risk of GDM and identify predictors applicable to the models.

Methods: Four reliable electronic databases were searched for studies that developed ML prediction models for GDM in the general population instead of among high-risk groups only. The novel Prediction Model Risk of Bias Assessment Tool (PROBAST) was used to assess the risk of bias of the ML models. The Meta-DiSc software program (version 1.4) was used to perform the meta-analysis and determination of heterogeneity. To limit the influence of heterogeneity, we also performed sensitivity analyses, a meta-regression, and subgroup analysis.

Results: A total of 25 studies that included women older than 18 years without a history of vital disease were analyzed. The pooled area under the receiver operating characteristic curve (AUROC) for ML models predicting GDM was 0.8492; the pooled sensitivity was 0.69 (95% CI 0.68-0.69; $P < .001$; $I^2 = 99.6\%$) and the pooled specificity was 0.75 (95% CI 0.75-0.75; $P < .001$; $I^2 = 100\%$). As one of the most commonly employed ML methods, logistic regression achieved an overall pooled AUROC of 0.8151, while non-logistic regression models performed better, with an overall pooled AUROC of 0.8891. Additionally, maternal age, family history of diabetes, BMI, and fasting blood glucose were the four most commonly used features of models established by the various feature selection methods.

Conclusions: Compared to current screening strategies, ML methods are attractive for predicting GDM. To expand their use, the importance of quality assessments and unified diagnostic criteria should be further emphasized.

(*J Med Internet Res* 2022;24(3):e26634) doi:[10.2196/26634](https://doi.org/10.2196/26634)

KEYWORDS

digital health; gestational diabetes mellitus; machine learning; prediction model; prognostic model

Introduction

According to the latest Global Diabetes Map (9th edition) released by the International Diabetes Federation, the number of patients with diabetes during pregnancy is increasing globally, with about 20.4 million (15.8%) women suffering from hyperglycemia; among them, 83.6% of cases were due to gestational diabetes mellitus (GDM) [1]. GDM, a common metabolic disease, is usually a transient disorder during pregnancy that resolves at delivery. Pregnant women with GDM are at greater risk of adverse pregnancy outcomes that threaten a normal birth. An oral glucose tolerance test (OGTT) is typically recommended to screen for GDM between the 24th and 28th weeks of gestation. Physicians usually measure the fasting plasma glucose concentration 1 to 2 hours after the patient ingests glucose [2]. The American Diabetes Association recommends that women be screened at the first prenatal examination to aid with the early identification of hyperglycemia risk. Nonetheless, GDM screening recommendations are controversial among international organizations regarding four aspects: (1) universal versus selective screening, (2) early pregnancy screening (ie, before pregnancy or at the first prenatal visit versus screening at 24-28 gestational weeks), (3) a one-step versus two-step approach, and (4) inconsistent diagnostic criteria (Table S1 in Multimedia Appendix 1) [3].

Machine learning (ML) methods have become favorable tools for disease prevention and management. For instance, the multivariate logistic regression (LR) model is a recognized ML algorithm for predicting diabetes and its complications. Furthermore, other methods, such as random forest (RF), extreme gradient boosting (XGBoost), and light gradient boosting machine (LightGBM), are also applied to diabetes-related problems. A growing number of studies have used such methods to identify risk factors of GDM and construct early prediction models for the disease [4,5]. ML presents a powerful tool for analyzing large amounts of diverse health care data and augmenting doctors' capabilities. However, ML has limitations that can lead to inaccurate predictions in some clinical scenarios, and the significance of its assessment was highlighted in a real-world study [6]. The US Food and Drug Administration (FDA) has issued guidance on software as a medical device that explains risk stratification and the analytical and clinical validation required of artificial intelligence (AI) tools in health care. IDx-DR, the first FDA-approved ML application to help make screening decisions, achieved high sensitivity (87%) and specificity (91%) for diabetic retinopathy in primary care clinics [7]. Most of the published prognostic models for GDM also showed acceptable discrimination and calibration [8], but they vary in quality and perform inconsistently.

Few systematic analyses of ML models for GDM are currently available. Here, we conducted a thorough meta-analysis of the predictive value of ML in GDM using a quality evaluation by the Prediction Model Risk of Bias Assessment Tool (PROBAST) and compared ML models with universal and selective screening methods. Essentially, we wondered if ML could be a new GDM screening option.

Methods

Research Design

This study was conducted according to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines (Table S2 in Multimedia Appendix 1) [9].

Search Methods

The PubMed, Web of Science, IEEE Xplore, and China National Knowledge Infrastructure databases were searched for articles published in English or Chinese between August 2019 and October 2020. We built up the search strategy according to the PICO (population, intervention, control, and outcomes) principle; for our study, "P" represents GDM populations, "I" represents ML methods as interventions, "C" represents gold standards as controls, and "O" represents the outcomes of prediction and diagnosis, such as sensitivity, specificity, and accuracy (Table S3 in Multimedia Appendix 1). The details of the search keywords are listed in Textbox S1 in Multimedia Appendix 1. Additionally, the reference list of each identified study was manually searched to identify any additional studies. NoteExpress 3.2 (Aegean) [10] and EndNote X7 (Clarivate) [11] were employed to manage the studies and remove duplicate items.

Inclusion and Exclusion Criteria

All studies included had to meet the following criteria: (1) published in English or Chinese; (2) included pregnant women from the general population, with a clear definition for GDM diagnosis; (3) included ML models for GDM prediction, with a clear description of the ML models; and (4) showed the performance of ML models, including sufficient data to enable the inference of sensitivity and specificity.

Articles in other languages, other types of articles (eg, reports and reviews), or those that used other measures for GDM detection were excluded. Four investigators (LY, WH, YW, and CG) participated in the literature screening to review all the studies that met the inclusion criteria. Each chosen article was screened at least twice, and disagreements were resolved by the reviewer (ZZ). Studies providing the most detailed information of variables and outcome indicators were kept for reference.

Data Extraction

Data extraction was performed independently by two investigators (LY and LZ) according to the existing literature and the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) standardized protocol [12]. A total of 25 studies were ultimately selected for the analysis. The following data were extracted from each study: (1) demographic information (ie, the country in which the data were gathered, the setting, the data source, the study design, the prediction temporality, and the outcome definition); (2) the data division method, the feature selection algorithms, the features of the model training, the ML prediction model type, and the model validation and application; (3) prediction outcomes, including accuracy, sensitivity, specificity,

and area under the receiver operating characteristic curve (AUROC); and (4) funding and ethics approval.

Quality and Bias Assessments

The PROBAST [13], which includes a total of 20 signaling questions in four domains (ie, participants, predictors, outcome, and analysis), was used as a tool for assessing the risk of bias and applicability of each included study.

Statistical Analysis

The performance of each ML model was described using the primary outcome measures of discrimination and calibration. Model discrimination or concordance index (C-index) is similar to the AUROC [14] and indicates its diagnostic or prognostic discrimination ability as none (AUROC ≤ 0.6), poor (AUROC > 0.6 to 0.7), fair (AUROC > 0.7 to 0.8), good (AUROC > 0.8 to 0.9), or optimum (AUROC > 0.9 to 1). Model calibration is a metric of goodness of fit that assesses the agreement between observed and predicted outcomes and reflects the stability of the model via calibration plots. The diagnostic odds ratio (DOR) was also calculated via the following equation:

$$\text{DOR} = \text{PLR} / \text{NLR} \text{ (1)}$$

where PLR is the positive likelihood ratio and NLR is the negative likelihood ratio. The PLR and NLR were calculated to express how frequently the model predicted GDM among the individuals with GDM versus among those without GDM using the following equations:

$$\text{PLR} = \text{Sensitivity} / (1 - \text{Specificity}) \text{ (2)}$$

$$\text{NLR} = (1 - \text{Sensitivity}) / \text{Specificity} \text{ (3)}$$

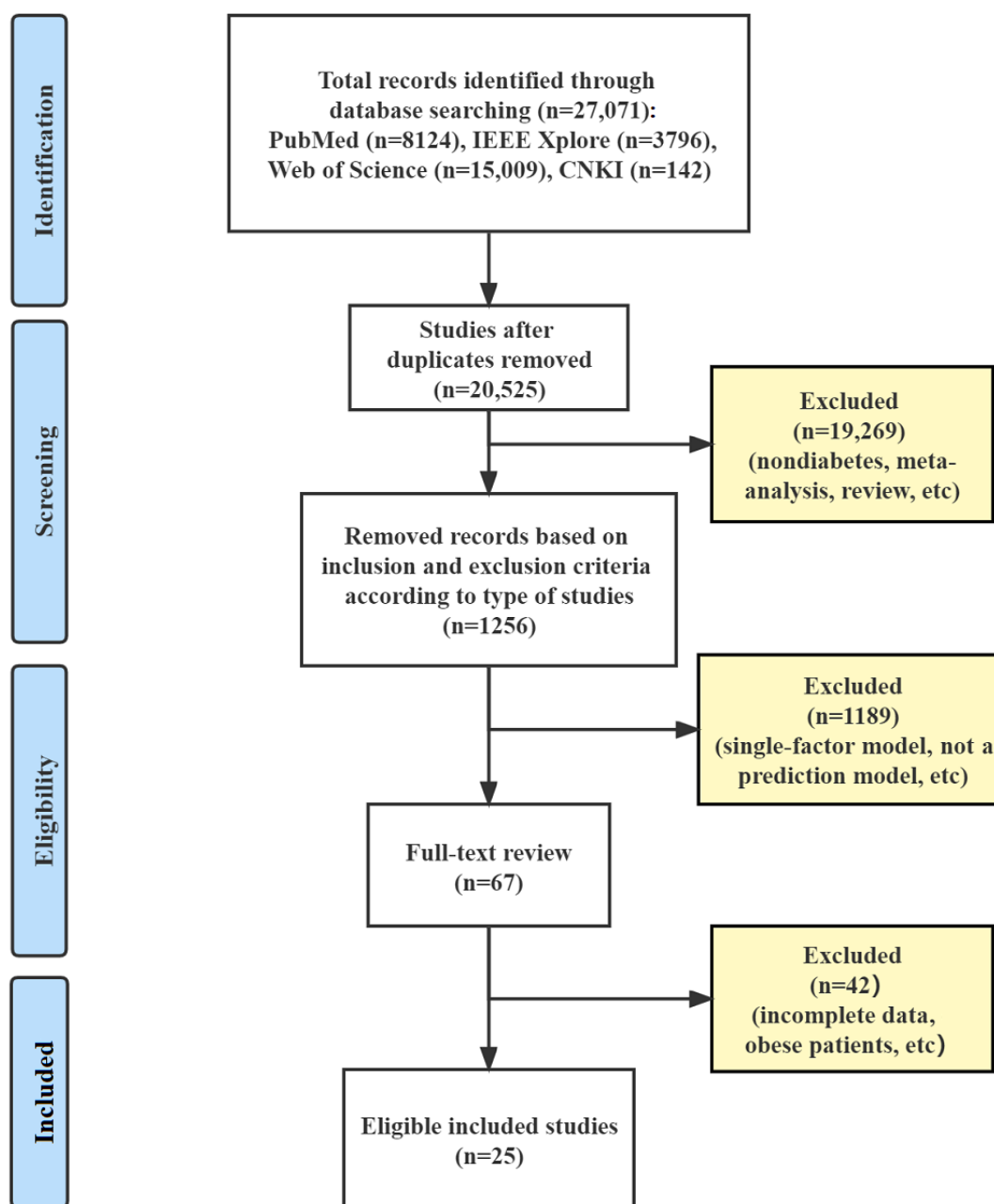
In this meta-analysis, the Meta-DiSc software program (version 1.4) [15] was used to calculate the pooled estimates of AUROC, sensitivity, specificity, PLR, NLR, and DOR. It was used to summarize the data from the included studies and graphically investigate the homogeneity among the studies. The I^2 test was used to assess the statistical heterogeneity among the included studies. An I^2 value of more than 75% indicated high heterogeneity among the studies [16]. The analysis of the included studies was divided into primary and subgroup analyses to judge the performances of the ML methods in predicting GDM in different clinical scenarios. Sensitivity analysis, subgroup analyses, and a meta-regression were also conducted to gain insight into potential sources of interstudy heterogeneity due to selector or inclusion criteria bias. The abilities of the different ML algorithms (eg, LR, Bayesian model, TreeNet, and GA-CatBoost [genetic algorithm category boosting]) for predicting GDM are discussed in the Subgroup Analysis section. The four predictive models with the highest and the lowest values were excluded from the sensitivity analysis to assess the impact of outliers on pooled sensitivity and specificity.

Results

Study Selection

A total of 27,071 studies were initially identified; of those, 1256 (4.6%) that discussed GDM were subjected to abstract screening. A total of 67 studies were subjected to full-text review; of those, 25 (37%) were included in the meta-analysis [17-33]. Figure 1 shows the PRISMA flow diagram of the study selection process.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for study selection. CNKI: China National Knowledge Infrastructure.



Study Characteristics

The articles' years of publication ranged from 2004 to 2020; 10 out of 25 (40%) were published in 2020 (Figure S1 in [Multimedia Appendix 1](#)). All studies included women older than 18 years without a history of heart or cerebrovascular disease or vital organ dysfunction. Out of 25 studies, 9 (36%) included patients with a history of GDM, while other studies excluded those cases with a history of GDM (Tables S4-S6 in [Multimedia Appendix 1](#)). The source data for ML training were mostly obtained from medical centers and maternity hospitals; some also included self-administered questionnaires. Out of 25 studies, 9 (36%) were conducted using data from a population-based prospective cohort or multicenter study. The sample size of the included studies varied from 134 to 66,687 participants.

Feature selection is an important step for ML training. Xiong et al [19] developed a prediction model for GDM risk in the first 19 weeks of gestation with several hepatic, renal, and coagulation function measures; they observed that a cutoff of prothrombin time and activated partial thromboplastin time could reliably predict GDM with a sensitivity of 88.3%, a specificity of 99.47%, and an AUROC of 94.2%. Maternal age, family history of diabetes, BMI, and fasting blood glucose were the four most commonly used features of the established models, whereas pregnancy-associated plasma protein A, leptin, lipocalin-2, adiponectin, weight gain, and soft drink intake during pregnancy were used in only one or two models each. [Table 1](#) [17-41] summarizes the most frequent features included in the prognostic models.

Table 1. The most frequent factors included in risk prediction models for gestational diabetes mellitus.

Study first author, year	Factors included in models											
	MA ^a (n=19)	FHD ^b (n=14)	BMI (n=12)	FPG ^c (n=11)	PBMI ^d (n=8)	HD ^e (n=8)	Ethnicity (n=6)	TG ^f (n=5)	HbA _{1c} ^g (n=4)	SBP ^h (n=3)	Height (n=3)	hsCRP ⁱ (n=3)
Gao, 2020 [21]	✓ ^j	✓	✓							✓	✓	
Liu, 2020 [17]	✓			✓	✓							
Miao, 2020 [28]		✓		✓			✓					
Tan, 2020 [41]	✓	✓		✓	✓							
Wu, 2020 [18]	✓	✓		✓		✓						
Xiong, 2020 [19]												
Ye, 2020 [20]	✓		✓	✓	✓	✓		✓	✓			
Zhang, 2020 [39]	✓	✓		✓	✓	✓			✓	✓		✓
Snyder, 2020 [40]	✓				✓		✓					
Cui, 2019 [25]	✓	✓			✓							
Zheng, 2019 [24]	✓		✓	✓				✓				
Nombo, 2018 [26]		✓	✓									
Sweeting, 2018 [27]		✓	✓			✓	✓	✓				
Xiao, 2018 [38]	✓		✓	✓				✓	✓		✓	
Huang, 2017 [23]	✓	✓		✓	✓							
Wu, 2017 [22]	✓		✓	✓							✓	✓
Gabbay-Benziv, 2015 [30]	✓		✓			✓	✓			✓		
Thériault, 2015 [29]	✓	✓	✓			✓	✓		✓			✓
Eleftheriades, 2014 [31]	✓											
Pintaudi, 2013 [32]		✓		✓	✓							
Savona-Ventura, 2013 [33]	✓			✓								
Tran, 2013 [34]	✓		✓									
Teede, 2011 [35]	✓	✓				✓	✓					
Vanleeuwen, 2009 [36]		✓	✓			✓	✓					
Caliskan, 2004 [37]	✓	✓	✓									

^aMA: maternal age.

^bFHD: family history of diabetes.

^cFPG: fasting plasma glucose.

^dPBMI: prepregnancy BMI.

^eHD: history of diabetes.

^fTG: triglyceride.

^gHbA_{1c}: hemoglobin A_{1c}.

^hSBP: systolic blood pressure.

ⁱhsCRP: high-sensitivity C-reaction protein.

^jA checkmark (✓) indicates that the factor was included.

The LR model was the most universally used model in the 25 studies (n=17, 68%) for predicting GDM risk, while 5 (20%) studies assessed the performance of other ML methods (ie, GA-CatBoost, XGBoost, Bayesian model, TreeNet, gradient-boosting decision tree [GBDT], adaptive boosting [AdaBoost], LightGBM, Vote, and RF). For measuring deep

learning performance, AUROC and the Youden index were most commonly used. AUROC was used in studies that did not provide the C-index. Out of 25 studies, 2 (8%) did not report metrics of model discrimination. Of the 25 studies, only 7 (28%) presented calibration measures. Internal validation was performed in 13 studies (52%) using random split or k-fold

cross-validation and bootstrapping. Only 4 studies out of 25 (16%) performed external validation.

Quality Assessment

Items from the PROBAST checklists (Multimedia Appendix 2) were used to assess the risk of bias and applicability of the prognostic prediction model studies. According to the criteria, the biases of participants in 4 out of 25 (16%) studies [30,31,33,36] were moderate, mainly due to debatable criteria, while biases in the other studies were low. Out of 25 studies, 24 (96%) study groups had a low bias of predictors, while 1 (4%) [32] had a moderate risk of bias because the prediction assessment was created with knowledge of the outcome data. The bias of outcome in 6 (24%) studies [22,30,33,35-37] was moderate due to the diagnostic criteria, while the others were low. A total of 8 (32%) groups had a moderate bias of analysis [21,23,27,31-33,36,37] and 1 (4%) [30] showed a high risk of bias due to an unreasonable number of participants with

outcomes. The overall bias rating of 10 (40%) groups [21-23,27,30,32,33,35-37] was moderate. Overall concerns regarding the applicability rating of 7 (28%) studies [21-23,27,29,32,33] were moderate because of excessive features in models making it difficult to collect data in actual use, whereas others were low (Table S7 in Multimedia Appendix 1).

Performance of ML Models for GDM Prediction

The overall pooled AUROC for ML models for predicting GDM was 0.8492 (Figure 2). Additional values were as follows: sensitivity 0.69 (95% CI 0.68-0.69; $P < .001$; $I^2 = 99.6%$; Figure 3); specificity 0.75 (95% CI 0.75-0.75; $P < .001$; $I^2 = 100%$; Figure 4); DOR 13.78 (95% CI 9.53-19.94; $P < .001$; $I^2 = 99.1%$); PLR 4.02 (95% CI 3.13-5.17; $P < .001$; $I^2 = 99.6%$); and NLR 0.31 (95% CI 0.26-0.38; $P < .001$; $I^2 = 98.7%$).

Figure 2. The overall pooled area under the receiver operating characteristic curve (AUROC) of machine learning models for gestational diabetes mellitus prediction. Q*: the sensitivity at the intersection of the SROC curve and the straight line (sensitivity=specificity); SROC: summary receiver operating characteristic.

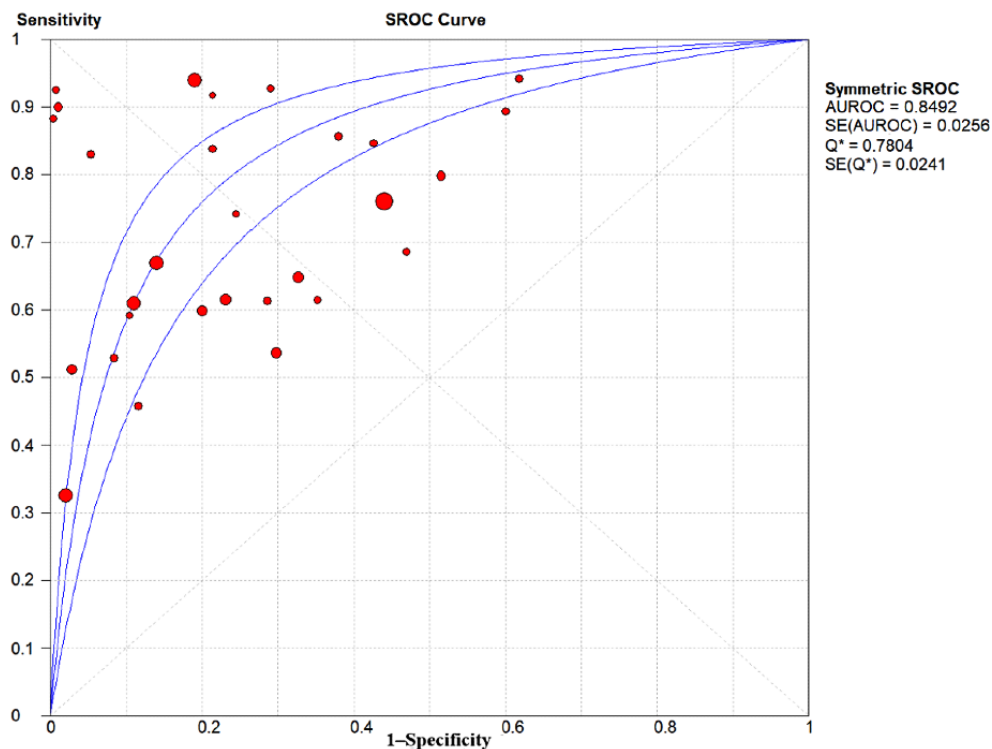


Figure 3. The overall pooled sensitivity of machine learning models for gestational diabetes mellitus prediction. First authors for each study are listed along the y-axis. The vertical red dotted lines are the 95% CIs of the pooled sensitivity. BY: Bayesian; DNN: deep neural network; GA-CB: GA-CatBoost (genetic algorithm category boosting); GBDT: gradient-boosting decision tree; KNN: k-nearest neighbors; LGB: LightGBM (light gradient boosting machine); LR: logistic regression; SVM: support vector machine; Tnet: TreeNet; XGB: XGBoost (extreme gradient boosting).

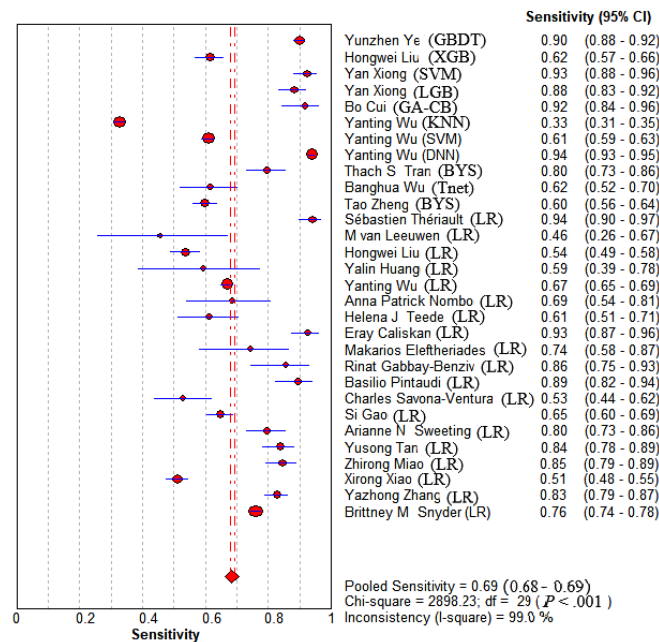
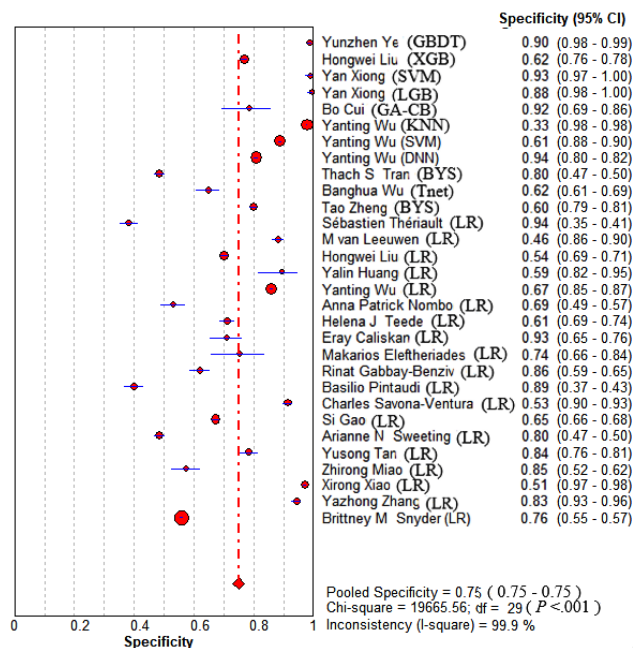


Figure 4. The overall pooled specificity of machine learning for gestational diabetes mellitus prediction. First authors for each study are listed along the y-axis. The vertical red dotted line is the 95% CI of the pooled specificity. BY: Bayesian; DNN: deep neural network; GA-CB: GA-CatBoost (genetic algorithm category boosting); GBDT: gradient-boosting decision tree; KNN: k-nearest neighbors; LGB: LightGBM (light gradient boosting machine); LR: logistic regression; SVM: support vector machine; Tnet: TreeNet; XGB: XGBoost (extreme gradient boosting).



Sensitivity Analysis

After the exclusion of the 4 (16%) models with the lowest and highest sensitivity and specificity, the random effects meta-analysis produced estimated pooled sensitivity of 0.73 (95% CI 0.72-0.74; P<.001; I²=98.3%) and pooled specificity of 0.73 (95% CI 0.72-0.73; P<.001; I²=99.8%). Therefore, the pooled estimates were deemed insensitive to the exclusion of outliers (Figure S2 in Multimedia Appendix 1).

Subgroup Analysis

The comparison of the GDM prediction performance results is shown in Table 2; forest plots are shown in Figures S3-S8 in Multimedia Appendix 1.

In this study, 19 prediction models were established using the LR models [17,18,20-23,26-33,35-39], and the overall pooled AUROC for the LR models for predicting GDM was 0.8151 (Figure 5). The overall pooled AUROC for non-LR models to predict GDM was 0.8891 (Figure 6), the highest value among

these subgroups. Further analysis of these non-LR methods showed that two support vector machine (SVM) models [19,20] achieved AUROC values of 0.82 and 0.98, respectively (Figure S9 in Multimedia Appendix 1), while two Bayesian models [24,34] achieved AUROC values of 0.766 and 0.71, respectively (Figure S10 in Multimedia Appendix 1). Interestingly, Ye et al [20] developed eight common ML methods—GBDT, AdaBoost, LightGBM, LR, Vote, XGBoost, decision tree (DT), and RF—and two common regression models to predict the occurrence of GDM with a data set of 822,242 patients. GBDT,

AdaBoost, and LightGBM (AUROC 0.70-0.75) were the top three models, while DT and RF were the worst models (AUROC 0.5-0.68) in that study. The capabilities of three ML methods were compared using data from 490 people [21]. The deep neural network model achieved the highest AUROC of 0.92, while the SVM and k-nearest neighbors (KNN) models achieved AUROC values of 0.82 and 0.68, respectively. XGBoost, LightGBM, GA-CatBoost, and TreeNet were used in 4 out of 25 (16%) studies and achieved AUROC values of 0.742, 0.942, 0.872, and 0.676, respectively [17,19,21,25] (Table 2).

Table 2. The comparison of performance of machine learning models in gestational diabetes mellitus (GDM) prediction applied to different subgroups.

Subgroup	Models (N=30), n (%)	AUROC ^a	Sensitivity (95% CI)	Specificity (95% CI)	PLR ^b (95% CI)	NLR ^c (95% CI)	DOR ^d (95% CI)
Overall	30 (100)	0.8492	0.69 (0.68-0.69)	0.75 (0.75-0.75)	4.02 (3.13-5.17)	0.31 (0.26-0.38)	13.78 (9.53-19.94)
0-13 weeks before diagnosis	16 (53)	0.8667	0.74 (0.73-0.75)	0.64 (0.64-0.64)	3.89 (2.92-5.19)	0.28 (0.22-0.36)	16.55 (9.52-28.77)
14-28 weeks before diagnosis	14 (47)	0.8365	0.64 (0.63-0.65)	0.85 (0.84-0.85)	3.90 (2.76-5.53)	0.35 (0.25-0.48)	11.67 (7.59-18.02)
With GDM history	11 (37)	0.8759	0.67 (0.66-0.68)	0.85 (0.85-0.86)	5.29 (3.39-8.25)	0.28 (0.18-0.44)	19.82 (11.49-34.13)
Without GDM history	19 (63)	0.8330	0.70 (0.66-0.68)	0.65 (0.64-0.65)	3.12 (2.52-3.86)	0.35 (0.30-0.41)	8.27 (5.14-13.29)
Logistic regression	19 (63)	0.8151	0.71 (0.70-0.72)	0.67 (0.67-0.67)	3.04 (2.37-3.89)	0.37 (0.32-0.43)	8.73 (5.99-12.73)
Non-logistic regression	11 (37)	0.8891	0.66 (0.65-0.67)	0.85 (0.85-0.86)	6.80 (4.45-10.37)	0.24 (0.15-0.38)	31.85 (15.93-63.69)

^aAUROC: area under receiver operating characteristic curve.

^bPLR: positive likelihood ratio.

^cNLR: negative likelihood ratio.

^dDOR: diagnostic odds ratio.

Figure 5. The overall pooled area under the receiver operating characteristic curve (AUROC) of logistic regression models for gestational diabetes mellitus prediction. Q*: the sensitivity at the intersection of the SROC curve and the straight line (sensitivity=specificity); SROC: summary receiver operating characteristic.

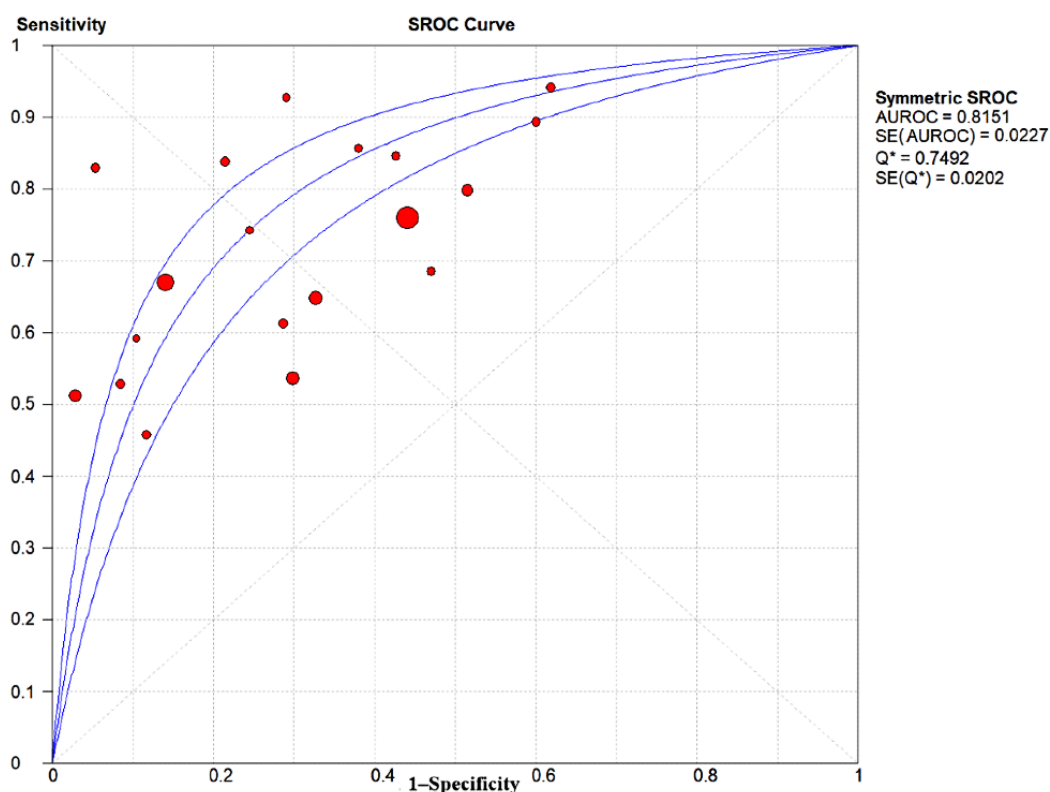
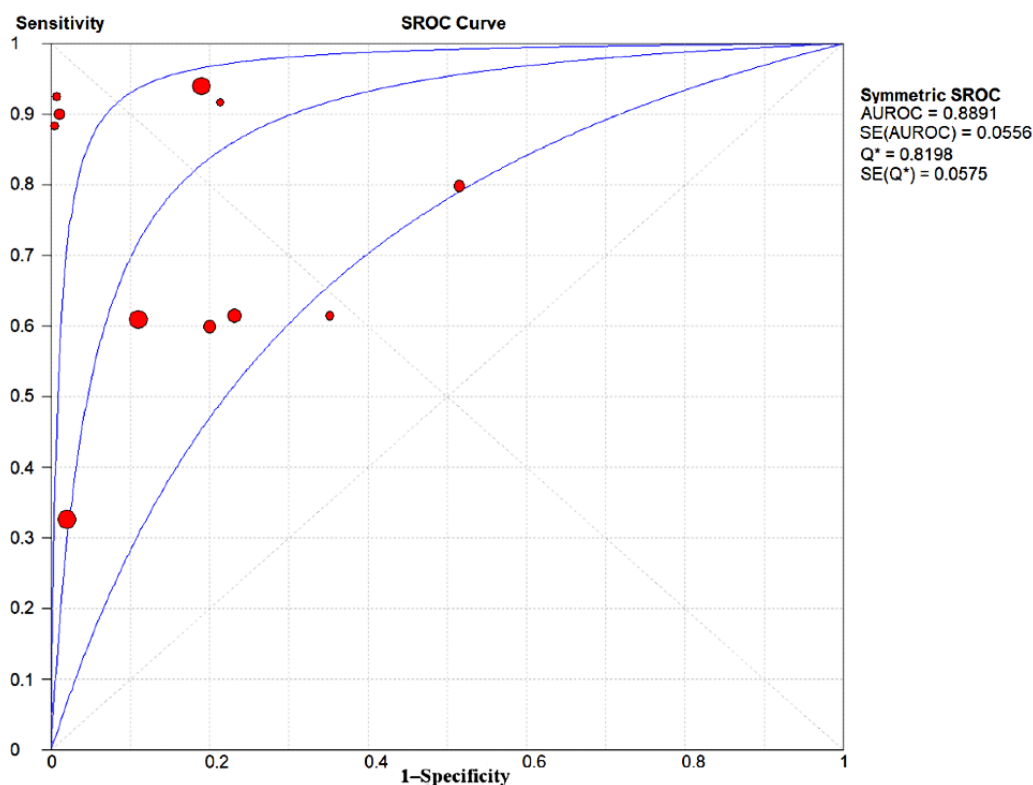


Figure 6. The overall pooled area under the receiver operating characteristic curve (AUROC) of non-logistic regression models for gestational diabetes mellitus prediction. Q^* : the sensitivity at the intersection of the SROC curve and the straight line (sensitivity=specificity); SROC: summary receiver operating characteristic.



Meta-regression

The meta-regression analysis was conducted due to the high level of interstudy heterogeneity [42]. Sample size, country where the data were collected, publication year, ML methods used, and model quality did not affect diagnostic accuracy ($P=0.13$). The antilogarithm transformations of the resulting estimated parameters could be interpreted as a relative DOR of the corresponding covariate, indicating the change in diagnostic performance of the test under study per unit increase in the covariate (Table S8 in [Multimedia Appendix 1](#)).

Discussion

Principal Findings

This study was a pilot meta-analysis evaluating the performance of ML models for predicting GDM. Its overall pooled estimation of 25 studies showed that ML models achieved high accuracy in early recognition of GDM patients. ML models could forecast based on data from 8 to 24 weeks' gestation. There was even a model that used prepregnancy features to predict the outcome up to 28 weeks in advance, suggesting the significance of ML models for GDM prediction. Compared to the census or existing screening methods, ML methods have certain advantages. Universal screening leads to 100% detection for physicians who usually make decisions based on an OGTT test, which may place an unnecessary burden on individual women and health care resources. Current selective screening strategies are based on a list of risk factors and have fixed sensitivity ($\pm 65\%$) and specificity ($\pm 80\%$). Although the ML methods do not provide greater benefit than current available screening strategies, an

advantage is that a preferred trade-off between sensitivity and specificity can be selected [43]. The choice of statistical method is more to compute a quantitative measure of existing data than to predict unknown data in a general and feasible way [44].

According to the subgroup analysis, models created using non-LR methods achieved the highest AUROC, suggesting that researchers should test more candidate models. One study aimed to review and compare the predictive performances of LR and other ML algorithms for developing or validating a multivariable prognostic prediction model for pregnancy care; that study also recommended a reanalysis of existing LR models for several pregnancy outcomes by comparing them with those algorithms that apply standard guidelines [45]. Among those non-LR models, ensemble methods, like LightGBM and GA-CatBoost, that are composed of multiple weaker models and are independently trained had a satisfactory result. Variables in the GBDT model underscored the advantage of identifying nonlinear relationships. The SVM model also achieved superior outcomes; that method builds a model that assigns new examples to one category or the other, making it a nonprobabilistic binary linear classifier. Methods like KNN, DT, and RF did not perform as well as the LightGBM and GA-CatBoost methods, which may be due to the fact that DT classifications are based on a single condition at the bottom, so small changes can lead to mistakes. For RF, the high dimension of medical data complicates the classification and prediction. Similarly, KNN cannot be used in high-dimensional feature spaces. Some researchers [23] found that the difference between two methods had no statistical significance, since LR models are suitable for simple data with linear relationships between variables and outcomes. Our study also found that LR models were conducive to achieving more

stable performance according to the summary receiver operating characteristic curve. The subgroup of 0 to 13 weeks before diagnosis achieved the highest pooled sensitivity, while the subgroup of 14 to 28 weeks before diagnosis achieved the highest specificity, meaning that ML may assist clinicians identify more patients in early screening and avoid excessive misdiagnosis in the second trimester.

The feature selection was also crucial for model performance and interpretation. Among the 25 studies, maternal age was used as a feature in 19 studies, as was previously reported and validated in our study. One of the included studies reported that the incidence of GDM increases after 25 years of age, the main reason being that the function of islet β -cells decreases with age, so the insulin antagonism of older adult pregnant women is aggravated [46]. Eight models considered GDM history to be a vital factor for predicting GDM. A DOR value of 21.09 appeared when a GDM history was included as a risk factor for predicting future GDM. Previous research discovered that women with GDM were more likely to have a family history of type 2 diabetes mellitus and a history of GDM, partially due to overlapping genetic bases between the diseases [18]. The nonsignificant association of GDM with a GDM history in other studies was a result of the overwhelming proportion of nulliparous women in their studies who had no risk of developing GDM. The association between GDM and blood lipid indexes, including triglyceride (TG), high-density lipoprotein, and low-density lipoprotein (LDL), has been studied, and TG level had the closest relationship with GDM [47]. Our research also found that although the levels of TG, total cholesterol, and LDL in the GDM group were higher than those in normal pregnant women in most included studies, only TG level was a high-risk factor of GDM after the feature selection. A novel model that included ultrasound data of maternal fat distribution and serum inflammatory factors observed that pregnant women with GDM had greater visceral fat thickness and subcutaneous fat thickness; the model also demonstrated that increased subcutaneous and visceral fat may lead to increased insulin resistance in muscle and adipose tissue [36]. Sweeting et al [27] observed higher leptin and lipocalin-2 levels and lower adiponectin levels in women who developed GDM and proposed adipokines as GDM features.

Strengths and Limitations

The main strength of the study is that its methodology was logical and described in sufficient detail to be reproducible. Almost all published prognostic models for GDM were included in this meta-analysis, which enabled their comparison. The data collection table was based on the characteristics of the GDM prediction models. Additionally, the novel PROBAST was used to assess the risk of bias and applicability of prognostic prediction model studies. The Quality Assessment of Diagnostic Accuracy Studies (QUADAS) tool is a widely used tool for estimating the bias and applicability of primary diagnostic accuracy studies, but is not perfectly suited for predictive models [48]. An increasing number of researchers prefer the PROBAST to the QUADAS tool for assessing the bias of AI-based models in systematic reviews as well as meta-analyses; this is the case because more details of the model, such as data source, processing, number of events per variable, feature selection,

model development, and model validation, were checked intensively [49-52]. We found that 14 development studies had a risk of bias in methodological quality or applicability, which may lead to overfitted prediction models. It is noteworthy that the quality of recent models is higher than that of those published earlier according to the PROBAST. Some bias could be prevented if the studies reported their research according to the TRIPOD initiative [12].

Despite our study's confirmation that ML models have promising prediction ability for GDM, there are some limitations to our research. The main limitations arose from the interstudy heterogeneity. First, the sample sizes and distributions differed among studies, affecting each model's performance and applicability. There were also a heterogenous variety of feature selection methods. Some researchers preferred the features that have a statistically significant association with GDM, while others included the factors based on existing knowledge from previously established models in combination with predictor reliability, consistency, applicability, availability, and cost. Second, the performance of a low-quality model might be overestimated when the analysis of the internal bias of the model is ignored. As some studies have bias to various degrees, the results of the studies in this analysis must be applied with caution. It should be noticed that the PROBAST is more likely to identify bias in prediction models than other tools designed for conventional diagnostic methods. The other limitation is that few models underwent external validation to test their extensibility. However, a previous study [8] performed an external validation of 12 published GDM prediction models and suggested that most of the published models showed acceptable discrimination and calibration, but the author pointed out possible heterogeneity in these models due to variations in GDM incidence in different populations.

Clinical Implications

Although several GDM scoring systems have been developed, none are widely recommended by current guidelines. Based on the discussion above, several items must be considered in order to maximize the advantages of ML models for predicting GDM in clinical practice for model researchers or for decision makers. For the former, we recommend that the decision concerning which feature selection methods and ML algorithms to use should be based on clinical need rather than accuracy. A model with excess features that are difficult to obtain in routine medicine is unlikely to be applied broadly. Researchers should also provide the process of data preprocessing and outcomes of validation, discrimination, calibration, and classification to elaborate the performance of models from multiple perspectives. For decision makers, we recommend that data sources, such as a population-based cohort designed for GDM research with a unified international diagnostic criterion, promote the ML methods in this target. Studies revealed that although electronic health records provide various data, including time series and images for novel ML methods, they have inherent biases that are influenced by the interaction of the patient with the health care system. In contrast, community-based predictions may robustly capture more asymptomatic high-risk cases [53]. The incidence of GDM based on the International Association of the Diabetes and Pregnancy Study Groups (IADPSG) (22.94%)

and the National Institute for Health and Care Excellence (21.72%) is over 3-fold higher than that based on the criteria from the 7th edition of the Chinese obstetrics and gynecology textbook (6.08%) published by the People's Medical Publishing House [54]. Some experts in China have advocated implementation of the IADPSG criteria because they believe that it will guide researchers to better understand the prevalence of GDM in different regions and ensure that the country's standards will be aligned with international ones. Nevertheless, researchers doubt that the IADPSG findings will apply to all populations, since those criteria were applied to mainly Caucasian women. All in all, it would indeed be helpful to unify the GDM diagnostic criteria as soon as possible. This

meta-analysis reported the advantages of ML models and the factors requiring attention. A similar meta-analysis of ML models and deep learning algorithms used to detect patients at risk of developing diabetes reported that AI-based automated tools provide substantial benefits for reducing screening costs and can replace earlier treatments [55].

Conclusions

In conclusion, ML methods demonstrate high performance and will be a more selective and cost-effective screening method for GDM. The importance of quality assessment and unified diagnostic criteria should be further emphasized.

Acknowledgments

This work was supported by grants from the National Key R&D Program of China (grants 2018YFC1314900 and 2018YFC1314902), Nantong "226 Project," Excellent Key Teachers in the "Qing Lan Project" of Jiangsu Colleges and Universities, Science and Technology Project of Nantong City (grant MS12020037), Jiangsu Students' Platform for Innovation and Entrepreneurship Training Program (grant 201910304108Y), and Jiangsu Postgraduate Research and Innovation Program (grant KYCX20_2836).

Authors' Contributions

HW supervised this study. LY, WH, YW, LZ, and CG performed literature retrieval and data extraction. KJ and YL provided technical support. ZZ performed experimental design and data processing and wrote the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary materials.

[DOCX File, 2194 KB - [jmir_v24i3e26634_app1.docx](#)]

Multimedia Appendix 2

Checklists from the Prediction Model Risk of Bias Assessment Tool (PROBAST).

[DOCX File, 22 KB - [jmir_v24i3e26634_app2.docx](#)]

References

1. Yuen L, Saedi P, Riaz M, Karuranga S, Divakar H, Levitt N, et al. Projections of the prevalence of hyperglycaemia in pregnancy in 2019 and beyond: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract* 2019 Nov;157:107841. [doi: [10.1016/j.diabres.2019.107841](#)] [Medline: [31518656](#)]
2. American Diabetes Association. Diagnosis and classification of diabetes mellitus. *Diabetes Care* 2013 Dec 19;37(Supplement_1):S81-S90. [doi: [10.2337/dc14-s081](#)]
3. Caissutti C, Berghella V. Scientific evidence for different options for GDM screening and management: Controversies and review of the literature. *Biomed Res Int* 2017;2017:2746471 [FREE Full text] [doi: [10.1155/2017/2746471](#)] [Medline: [28497042](#)]
4. Wang J. Prediction of Gestational Diabetes Based on Machine Learning [master's thesis; document in Chinese]. Chengdu, China: Southwestern University of Finance and Economics; 2019. URL: <https://tinyurl.com/yck82cvx>
5. Artzi NS, Shilo S, Hadar E, Rossman H, Barbash-Hazan S, Ben-Haroush A, et al. Prediction of gestational diabetes based on nationwide electronic health records. *Nat Med* 2020 Jan;26(1):71-76. [doi: [10.1038/s41591-019-0724-8](#)] [Medline: [31932807](#)]
6. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. *Lancet Oncol* 2019 May;20(5):e262-e273. [doi: [10.1016/S1470-2045\(19\)30149-4](#)] [Medline: [31044724](#)]
7. Changes to Existing Medical Software Policies Resulting From Section 3060 of the 21st Century Cures Act: Guidance for Industry and Food and Drug Administration Staff. Rockville, MD: US Food and Drug Administration; 2019 Sep. URL: <https://tinyurl.com/yy8zuajc>

8. Lamain-de Ruyter M, Kwee A, Naaktgeboren CA, de Groot I, Evers IM, Groenendaal F, et al. External validation of prognostic models to predict risk of gestational diabetes mellitus in one Dutch cohort: Prospective multicentre cohort study. *BMJ* 2016 Aug 30;354:i4338 [FREE Full text] [doi: [10.1136/bmj.i4338](https://doi.org/10.1136/bmj.i4338)] [Medline: [27576867](https://pubmed.ncbi.nlm.nih.gov/27576867/)]
9. 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](https://doi.org/10.1371/journal.pmed.1000097)] [Medline: [19621072](https://pubmed.ncbi.nlm.nih.gov/19621072/)]
10. NoteExpress. URL: <http://www.inoteexpress.com/wiki/index.php/%E9%A6%96%E9%A1%B5> [accessed 2020-11-01]
11. Eapen B. EndNote 7.0. *Indian J Dermatol Venereol Leprol* 2006;72(2):165-166. [doi: [10.4103/0378-6323.25654](https://doi.org/10.4103/0378-6323.25654)] [Medline: [16707834](https://pubmed.ncbi.nlm.nih.gov/16707834/)]
12. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD Statement. *Eur J Clin Invest* 2015 Feb;45(2):204-214. [doi: [10.1111/eci.12376](https://doi.org/10.1111/eci.12376)] [Medline: [25623047](https://pubmed.ncbi.nlm.nih.gov/25623047/)]
13. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: A tool to assess risk of bias and applicability of prediction model studies: Explanation and elaboration. *Ann Intern Med* 2019 Jan 01;170(1):W1. [doi: [10.7326/m18-1377](https://doi.org/10.7326/m18-1377)]
14. Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models. *Epidemiology* 2010;21(1):128-138. [doi: [10.1097/ede.0b013e3181c30fb2](https://doi.org/10.1097/ede.0b013e3181c30fb2)]
15. Zamora J, Abaira V, Muriel A, Khan K, Coomarasamy A. Meta-DiSc: A software for meta-analysis of test accuracy data. *BMC Med Res Methodol* 2006 Jul 12;6:31 [FREE Full text] [doi: [10.1186/1471-2288-6-31](https://doi.org/10.1186/1471-2288-6-31)] [Medline: [16836745](https://pubmed.ncbi.nlm.nih.gov/16836745/)]
16. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: [10.1136/bmj.327.7414.557](https://doi.org/10.1136/bmj.327.7414.557)] [Medline: [12958120](https://pubmed.ncbi.nlm.nih.gov/12958120/)]
17. Liu H, Li J, Leng J, Wang H, Liu J, Li W, et al. Machine learning risk score for prediction of gestational diabetes in early pregnancy in Tianjin, China. *Diabetes Metab Res Rev* 2021 Jul 09;37(5):e3397. [doi: [10.1002/dmrr.3397](https://doi.org/10.1002/dmrr.3397)] [Medline: [32845061](https://pubmed.ncbi.nlm.nih.gov/32845061/)]
18. Wu Y, Zhang C, Mol BW, Kawai A, Li C, Chen L, et al. Early prediction of gestational diabetes mellitus in the Chinese population via advanced machine learning. *J Clin Endocrinol Metab* 2021 Mar 08;106(3):e1191-e1205 [FREE Full text] [doi: [10.1210/clinem/dgaa899](https://doi.org/10.1210/clinem/dgaa899)] [Medline: [33351102](https://pubmed.ncbi.nlm.nih.gov/33351102/)]
19. Xiong Y, Lin L, Chen Y, Salerno S, Li Y, Zeng X, et al. Prediction of gestational diabetes mellitus in the first 19 weeks of pregnancy using machine learning techniques. *J Matern Fetal Neonatal Med* 2020 Aug 06:1-7. [doi: [10.1080/14767058.2020.1786517](https://doi.org/10.1080/14767058.2020.1786517)] [Medline: [32762275](https://pubmed.ncbi.nlm.nih.gov/32762275/)]
20. Ye Y, Xiong Y, Zhou Q, Wu J, Li X, Xiao X. Comparison of machine learning methods and conventional logistic regressions for predicting gestational diabetes using routine clinical data: A retrospective cohort study. *J Diabetes Res* 2020;2020:4168340 [FREE Full text] [doi: [10.1155/2020/4168340](https://doi.org/10.1155/2020/4168340)] [Medline: [32626780](https://pubmed.ncbi.nlm.nih.gov/32626780/)]
21. Gao S, Leng J, Liu H, Wang S, Li W, Wang Y, et al. Development and validation of an early pregnancy risk score for the prediction of gestational diabetes mellitus in Chinese pregnant women. *BMJ Open Diabetes Res Care* 2020 Apr;8(1):e000909 [FREE Full text] [doi: [10.1136/bmjdr-2019-000909](https://doi.org/10.1136/bmjdr-2019-000909)] [Medline: [32327440](https://pubmed.ncbi.nlm.nih.gov/32327440/)]
22. Wu B, Huang H, Yao Q, Deng R, Li H, Liang T. Application of big data and artificial intelligence methods in the prediction of gestational diabetes mellitus. *Chin J Health Inf Manag* 2017;14(6):832-837. [doi: [10.3969/j.issn.1672-5166.2017.06.016](https://doi.org/10.3969/j.issn.1672-5166.2017.06.016)]
23. Huang Y. Construction of a Gestational Diabetes Risk Model Based on the Guangxi Birth Cohort and Analysis of its Impact on Pregnancy Outcomes [master's thesis; document in Chinese]. Guangxi, China: Guangxi Medical University; 2017. URL: <https://tinyurl.com/2p8tredx> [accessed 2020-09-14]
24. Zheng T, Ye W, Wang X, Li X, Zhang J, Little J, et al. A simple model to predict risk of gestational diabetes mellitus from 8 to 20 weeks of gestation in Chinese women. *BMC Pregnancy Childbirth* 2019 Jul 19;19(1):252 [FREE Full text] [doi: [10.1186/s12884-019-2374-8](https://doi.org/10.1186/s12884-019-2374-8)] [Medline: [31324151](https://pubmed.ncbi.nlm.nih.gov/31324151/)]
25. Cui B. Study of Cascade GA-CatBoost in Predictive Diagnosis of Gestational Diabetes Mellitus [master's thesis; document in Chinese]. Taiyuan, China: Taiyuan University of Technology; 2019. URL: <http://150.138.141.24/kcms/detail/detail.aspx?filename=1019872456.nh&dbcode=CMFD&dbname=CMFDREF> [accessed 2022-01-25]
26. Nombo AP, Mwanri AW, Brouwer-Brolsma EM, Ramaiya KL, Feskens EJ. Gestational diabetes mellitus risk score: A practical tool to predict gestational diabetes mellitus risk in Tanzania. *Diabetes Res Clin Pract* 2018 Nov;145:130-137. [doi: [10.1016/j.diabres.2018.05.001](https://doi.org/10.1016/j.diabres.2018.05.001)] [Medline: [29852237](https://pubmed.ncbi.nlm.nih.gov/29852237/)]
27. Sweeting A, Wong J, Appelblom H, Ross G, Kouru H, Williams P, et al. A novel early pregnancy risk prediction model for gestational diabetes mellitus. *Fetal Diagn Ther* 2019;45(2):76-84. [doi: [10.1159/000486853](https://doi.org/10.1159/000486853)] [Medline: [29898442](https://pubmed.ncbi.nlm.nih.gov/29898442/)]
28. Miao Z, Wu H, Zhang Y, Gao Y, Jiang Y, Zhao Y, et al. The values of early pregnancy glucose and lipid metabolism indexes in predicting gestational diabetes mellitus with different prepregnancy body mass index. *Chin J Diabetes* 2020;12(7):462-468. [doi: [10.3760/cma.j.cn115791-20200401-00189](https://doi.org/10.3760/cma.j.cn115791-20200401-00189)]
29. Thériault S, Giguère Y, Massé J, Girouard J, Forest J. Early prediction of gestational diabetes: A practical model combining clinical and biochemical markers. *Clin Chem Lab Med* 2016 Mar;54(3):509-518. [doi: [10.1515/cclm-2015-0537](https://doi.org/10.1515/cclm-2015-0537)] [Medline: [26351946](https://pubmed.ncbi.nlm.nih.gov/26351946/)]

30. Gabbay-Benziv R, Doyle L, Blitzer M, Baschat A. First trimester prediction of maternal glycemic status. *J Perinat Med* 2015 May;43(3):283-289. [doi: [10.1515/jpm-2014-0149](https://doi.org/10.1515/jpm-2014-0149)] [Medline: [25153547](https://pubmed.ncbi.nlm.nih.gov/25153547/)]
31. Eleftheriades M, Papastefanou I, Lambrinouadaki I, Kappou D, Lavranos D, Akalestos A, et al. Elevated placental growth factor concentrations at 11-14 weeks of gestation to predict gestational diabetes mellitus. *Metabolism* 2014 Nov;63(11):1419-1425. [doi: [10.1016/j.metabol.2014.07.016](https://doi.org/10.1016/j.metabol.2014.07.016)] [Medline: [25173717](https://pubmed.ncbi.nlm.nih.gov/25173717/)]
32. Pintaudi B, Di Vieste G, Corrado F, Lucisano G, Pellegrini F, Giunta L, et al. Improvement of selective screening strategy for gestational diabetes through a more accurate definition of high-risk groups. *Eur J Endocrinol* 2013 Nov 22;170(1):87-93 [FREE Full text] [doi: [10.1530/eje-13-0759](https://doi.org/10.1530/eje-13-0759)]
33. Savona-Ventura C, Vassallo J, Marre M, Karamanos BG, MGSD-GDM Study Group. A composite risk assessment model to screen for gestational diabetes mellitus among Mediterranean women. *Int J Gynaecol Obstet* 2013 Mar;120(3):240-244. [doi: [10.1016/j.ijgo.2012.10.016](https://doi.org/10.1016/j.ijgo.2012.10.016)] [Medline: [23279935](https://pubmed.ncbi.nlm.nih.gov/23279935/)]
34. Tran TS, Hirst JE, Do MAT, Morris JM, Jeffery HE. Early prediction of gestational diabetes mellitus in Vietnam: Clinical impact of currently recommended diagnostic criteria. *Diabetes Care* 2013 Mar;36(3):618-624 [FREE Full text] [doi: [10.2337/dc12-1418](https://doi.org/10.2337/dc12-1418)] [Medline: [23160727](https://pubmed.ncbi.nlm.nih.gov/23160727/)]
35. Teede H, Harrison C, Teh W, Paul E, Allan C. Gestational diabetes: Development of an early risk prediction tool to facilitate opportunities for prevention. *Aust N Z J Obstet Gynaecol* 2011;51(6):499-504. [doi: [10.1111/j.1479-828x.2011.01356.x](https://doi.org/10.1111/j.1479-828x.2011.01356.x)]
36. van Leeuwen M, Opmeer B, Zweers E, van Ballegooie E, ter Brugge HG, de Valk HW, et al. Estimating the risk of gestational diabetes mellitus: A clinical prediction model based on patient characteristics and medical history. *BJOG* 2010 Jan;117(1):69-75. [doi: [10.1111/j.1471-0528.2009.02425.x](https://doi.org/10.1111/j.1471-0528.2009.02425.x)] [Medline: [20002371](https://pubmed.ncbi.nlm.nih.gov/20002371/)]
37. Caliskan E, Kayikioglu F, Oztürk N, Koc S, Haberal A. A population-based risk factor scoring will decrease unnecessary testing for the diagnosis of gestational diabetes mellitus. *Acta Obstet Gynecol Scand* 2004 Jun;83(6):524-530 [FREE Full text] [doi: [10.1111/j.0001-6349.2004.00389.x](https://doi.org/10.1111/j.0001-6349.2004.00389.x)] [Medline: [15144332](https://pubmed.ncbi.nlm.nih.gov/15144332/)]
38. Xiao X, Dong X, Li X, Yin C. The early predictive value of the basic characteristics of pregnant women and 14-20 weeks of glucose and lipid metabolism indicators for gestational diabetes. *Chin J Pract Gynecol Obstet* 2018;34(04):429-432. [doi: [10.19538/j.fk2018040120](https://doi.org/10.19538/j.fk2018040120)]
39. Zhang Y, Zhou L, Tian L, Li X, Zhang G, Qin J, et al. A mid-pregnancy risk prediction model for gestational diabetes mellitus based on the maternal status in combination with ultrasound and serological findings. *Exp Ther Med* 2020 Jul;20(1):293-300 [FREE Full text] [doi: [10.3892/etm.2020.8690](https://doi.org/10.3892/etm.2020.8690)] [Medline: [32536997](https://pubmed.ncbi.nlm.nih.gov/32536997/)]
40. Snyder BM, Baer RJ, Oltman SP, Robinson JG, Breheny PJ, Saftlas AF, et al. Early pregnancy prediction of gestational diabetes mellitus risk using prenatal screening biomarkers in nulliparous women. *Diabetes Res Clin Pract* 2020 May;163:108139 [FREE Full text] [doi: [10.1016/j.diabres.2020.108139](https://doi.org/10.1016/j.diabres.2020.108139)] [Medline: [32272192](https://pubmed.ncbi.nlm.nih.gov/32272192/)]
41. Tan Y. Clinical research on gestational diabetes models in Changchun, China. *Matern Child Health Care China* 2020;35(16):2957-2959. [doi: [10.19829/j.zgfybj.issn.1001-4411.2020.16.007](https://doi.org/10.19829/j.zgfybj.issn.1001-4411.2020.16.007)]
42. Lijmer JG, Bossuyt PMM, Heisterkamp SH. Exploring sources of heterogeneity in systematic reviews of diagnostic tests. *Stat Med* 2002 Jun 15;21(11):1525-1537. [doi: [10.1002/sim.1185](https://doi.org/10.1002/sim.1185)] [Medline: [12111918](https://pubmed.ncbi.nlm.nih.gov/12111918/)]
43. Meertens LJE, Scheepers HCJ, van Kuijk SMJ, Roeleveld N, Aardenburg R, van Dooren IMA, et al. External validation and clinical utility of prognostic prediction models for gestational diabetes mellitus: A prospective cohort study. *Acta Obstet Gynecol Scand* 2020 Jul;99(7):891-900 [FREE Full text] [doi: [10.1111/aogs.13811](https://doi.org/10.1111/aogs.13811)] [Medline: [31955406](https://pubmed.ncbi.nlm.nih.gov/31955406/)]
44. Bzdok D, Altman N, Krzywinski M. Statistics versus machine learning. *Nat Methods* 2018 Apr;15(4):233-234 [FREE Full text] [doi: [10.1038/nmeth.4642](https://doi.org/10.1038/nmeth.4642)] [Medline: [30100822](https://pubmed.ncbi.nlm.nih.gov/30100822/)]
45. Sufriyana H, Husnayain A, Chen Y, Kuo C, Singh O, Yeh T, et al. Comparison of multivariable logistic regression and other machine learning algorithms for prognostic prediction studies in pregnancy care: Systematic review and meta-analysis. *JMIR Med Inform* 2020 Nov 17;8(11):e16503 [FREE Full text] [doi: [10.2196/16503](https://doi.org/10.2196/16503)] [Medline: [33200995](https://pubmed.ncbi.nlm.nih.gov/33200995/)]
46. Liu Y, Du L. The influence of aging factors on pregnancy outcome [journal in Chinese]. *Clinical Medical Engineering* 2011;18(3):365-366 [FREE Full text] [doi: [10.3969/j.issn.1674-4659.2011.03.0365](https://doi.org/10.3969/j.issn.1674-4659.2011.03.0365)]
47. Niromanesh S, Shirazi M, Dastgerdy E, Sharbaf FR, Shirazi M, Khazaeipour Z. Association of hypertriglyceridaemia with pre-eclampsia, preterm birth, gestational diabetes and uterine artery pulsatility index. *Natl Med J India* 2012;25(5):265-267. [Medline: [23448624](https://pubmed.ncbi.nlm.nih.gov/23448624/)]
48. Whiting PF, Rutjes AWS, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: A revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011 Oct 18;155(8):529-536 [FREE Full text] [doi: [10.7326/0003-4819-155-8-201110180-00009](https://doi.org/10.7326/0003-4819-155-8-201110180-00009)] [Medline: [22007046](https://pubmed.ncbi.nlm.nih.gov/22007046/)]
49. Palazón-Bru A, Mares-García E, López-Bru D, Mares-Arambul E, Folgado-de la Rosa DM, de Los Ángeles Carbonell-Torregrosa M, et al. A critical appraisal of the clinical applicability and risk of bias of the predictive models for mortality and recurrence in patients with oropharyngeal cancer: Systematic review. *Head Neck* 2020 Apr;42(4):763-773. [doi: [10.1002/hed.26025](https://doi.org/10.1002/hed.26025)] [Medline: [31762119](https://pubmed.ncbi.nlm.nih.gov/31762119/)]
50. Di Tanna G, Wirtz H, Burrows K, Globe G. Evaluating risk prediction models for adults with heart failure: A systematic literature review. *PLoS One* 2020;15(1):e0224135 [FREE Full text] [doi: [10.1371/journal.pone.0224135](https://doi.org/10.1371/journal.pone.0224135)] [Medline: [31940350](https://pubmed.ncbi.nlm.nih.gov/31940350/)]

51. Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies. *BMJ* 2020 Mar 25;368:m689 [FREE Full text] [doi: [10.1136/bmj.m689](https://doi.org/10.1136/bmj.m689)] [Medline: [32213531](https://pubmed.ncbi.nlm.nih.gov/32213531/)]
52. Dretzke J, Chuchu N, Agarwal R, Herd C, Chua W, Fabritz L, et al. Predicting recurrent atrial fibrillation after catheter ablation: A systematic review of prognostic models. *Europace* 2020 May 01;22(5):748-760 [FREE Full text] [doi: [10.1093/europace/euaa041](https://doi.org/10.1093/europace/euaa041)] [Medline: [32227238](https://pubmed.ncbi.nlm.nih.gov/32227238/)]
53. Silva KD, Lee WK, Forbes A, Demmer RT, Barton C, Enticott J. Use and performance of machine learning models for type 2 diabetes prediction in community settings: A systematic review and meta-analysis. *Int J Med Inform* 2020 Nov;143:104268. [doi: [10.1016/j.ijmedinf.2020.104268](https://doi.org/10.1016/j.ijmedinf.2020.104268)] [Medline: [32950874](https://pubmed.ncbi.nlm.nih.gov/32950874/)]
54. He Z, Xie H, Liang S, Tang Y, Ding W, Wu Y, et al. Influence of different diagnostic criteria on gestational diabetes mellitus incidence and medical expenditures in China. *J Diabetes Investig* 2019 Sep;10(5):1347-1357 [FREE Full text] [doi: [10.1111/jdi.13008](https://doi.org/10.1111/jdi.13008)] [Medline: [30663279](https://pubmed.ncbi.nlm.nih.gov/30663279/)]
55. Islam MM, Yang H, Poly TN, Jian W, Li YCJ. Deep learning algorithms for detection of diabetic retinopathy in retinal fundus photographs: A systematic review and meta-analysis. *Comput Methods Programs Biomed* 2020 Jul;191:105320. [doi: [10.1016/j.cmpb.2020.105320](https://doi.org/10.1016/j.cmpb.2020.105320)] [Medline: [32088490](https://pubmed.ncbi.nlm.nih.gov/32088490/)]

Abbreviations

AdaBoost: adaptive boosting
AI: artificial intelligence
AUROC: area under the receiver operating characteristic curve
C-index: concordance index
DOR: diagnostic odds ratio
DT: decision tree
FDA: Food and Drug Administration
GA-CatBoost: genetic algorithm category boosting
GBDT: gradient-boosting decision tree
GDM: gestational diabetes mellitus
IADPSG: International Association of the Diabetes and Pregnancy Study Groups
KNN: k-nearest neighbors
LDL: low-density lipoprotein
LightGBM: light gradient boosting machine
LR: logistic regression
ML: machine learning
NLR: negative likelihood ratio
OGTT: oral glucose tolerance test
PICO: population, intervention, control, and outcomes
PLR: positive likelihood ratio
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST: Prediction Model Risk of Bias Assessment Tool
QUADAS: Quality Assessment of Diagnostic Accuracy Studies
RF: random forest
SVM: support vector machine
TG: triglyceride
TRIPOD: Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
XGBoost: extreme gradient boosting

Edited by G Eysenbach; submitted 19.12.20; peer-reviewed by E van der Velde, A McLean, E Da Silva, AUR Bacha, Y Mao; comments to author 16.01.21; revised version received 11.03.21; accepted 10.12.21; published 16.03.22.

Please cite as:

Zhang Z, Yang L, Han W, Wu Y, Zhang L, Gao C, Jiang K, Liu Y, Wu H
Machine Learning Prediction Models for Gestational Diabetes Mellitus: Meta-analysis
J Med Internet Res 2022;24(3):e26634
URL: <https://www.jmir.org/2022/3/e26634>
doi: [10.2196/26634](https://doi.org/10.2196/26634)
PMID: [35294369](https://pubmed.ncbi.nlm.nih.gov/35294369/)

©Zheqing Zhang, Luqian Yang, Wentao Han, Yaoyu Wu, Linhui Zhang, Chun Gao, Kui Jiang, Yun Liu, Huiqun Wu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 16.03.2022. 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.

Original Paper

The Mediating Role of Patients' Trust Between Web-Based Health Information Seeking and Patients' Uncertainty in China: Cross-sectional Web-Based Survey

Wei Dong^{1,2*}, PhD; Xiangxi Lei^{3*}, MPhil; Yongmei Liu¹, PhD

¹School of Business, Central South University, Changsha, China

²Department of Information Systems, City University of Hong Kong, Hong Kong, Hong Kong

³China Mobile Group Hunan Company Limited, Changsha, China

*these authors contributed equally

Corresponding Author:

Yongmei Liu, PhD

School of Business

Central South University

932 Lushan South Road

Changsha, 410083

China

Phone: 86 13974834821

Email: liuyongmeicn@163.com

Abstract

Background: In the physician-patient relationship, patients' uncertainty about diseases and the lack of trust in physicians not only hinder patients' rehabilitation but also disrupt the harmony in this relationship. With the development of the web-based health industry, patients can easily access web-based information about health care and physicians, thus reducing patients' uncertainty to some extent. However, it is not clear how patients' web-based health information-seeking behaviors reduce their uncertainty.

Objective: On the basis of the principal-agent theory and the perspective of uncertainty reduction, this study aims to investigate the mechanism of how web-based disease-related information and web-based physician-related information reduce patients' uncertainty.

Methods: A web-based survey involving 337 participants was conducted. In this study, we constructed a structural equation model and used SmartPLS (version 3.3.3; SmartPLS GmbH) software to test the reliability and validity of the measurement model. The path coefficients of the structural model were also calculated to test our hypotheses.

Results: By classifying patients' uncertainties into those concerning diseases and those concerning physicians, this study identified the different roles of the two types of patients' uncertainty and revealed that web-based disease-related information quality and web-based physician-related information can act as uncertainty mitigators. The quality of disease-related information reduces patients' perceived information scarcity about the disease ($\beta=-.588$; $P<.001$), and the higher the information scarcity perceived by patients, the higher their uncertainty toward the disease ($\beta=.111$; $P=.02$). As for physician-related information, web-based word-of-mouth information about physicians reduces patients' perceived information scarcity about the physician ($\beta=-.511$; $P<.001$), mitigates patients' fears about physician opportunism ($\beta=-.268$; $P<.001$), and facilitates patients' trust ($\beta=.318$; $P<.001$). These factors further influence patients' uncertainty about the physician. In addition, from the test of mediating effect, patients' trust in the physician fully mediates the relationship between their perceived information scarcity about the physician's medical service and their uncertainty about the physician. Patients' trust also partially mediates the relationship between their fear of the physician's opportunism and their uncertainty about the physician. As for the two different types of uncertainty, patients' uncertainty about the physician also increases their uncertainty about the diseases ($\beta=.587$; $P<.001$).

Conclusions: This study affirms the role of disease-related web-based information quality and physician-related web-based word-of-mouth information in reducing patients' uncertainties. With regard to the traits of principal-agent relationships, this study describes the influence mechanism based on patients' perceived information scarcity, fears of physicians' opportunism, and patients' trust. Moreover, information about physicians is effective in reducing patients' uncertainties, but only if the information

enhances patients' trust in their physicians. This research generates new insights into understanding the impact of web-based health information on patients' uncertainties.

(*J Med Internet Res* 2022;24(3):e25275) doi:[10.2196/25275](https://doi.org/10.2196/25275)

KEYWORDS

patient trust; online health information quality; online word-of-mouth; patient uncertainty; principal-agent theory; physician-patient relationship

Introduction

Background

In a physician-patient relationship, it is always difficult for patients to evaluate medical services and their physicians because medical services are typical credence products [1,2]. Patients also lack the specialized knowledge to judge whether a physician's treatments would be helpful before the treatments begin. Therefore, as principals, in this typical principal-agent relationship, patients face many uncertainties.

Previous research has explored how to achieve better treatment outcomes by reducing patients' uncertainty [3-6]. The uncertainty in this principal-agent relationship is caused by information problems [7], such as hiding information and hiding behaviors; therefore, to reduce patients' uncertainties, it is important to provide patients with more information. With the rapid development of patient-centered care [8], the physician-patient relationship is gradually changing from the traditional physician-led model to a new type of patient-centered diagnosis and treatment, with increasing emphasis on the role of patients [9]. The role of patients is changing from passive information recipients to active participants in medical decision-making [10]. The development of the eHealth industry has led to an increase in the number of patients who become electronic patients, namely, e-patients [9]. The channels for e-patients to obtain information about diseases and physicians have expanded, and this information can enhance patient-centered care [8]. For example, in a survey by Wong and Cheung [11], 97.32% (1162/1194) of the respondents used the internet, of which 87.44% (1016/1162) had used the internet to find health information. In a survey by Hedges and Couey [12], 90% of patients used web-based reviews to evaluate their physicians. By actively acquiring information about diseases and physicians through electronic information technology, e-patients can enhance their understanding of their medical condition and have a sense of control over their health, while reducing their uncertainties about the consultation processes and the physicians.

Although web-based information can reduce patients' uncertainty to some extent, information overload can pose a major challenge [13], leading to confusion in e-patients. Incorrect information does not effectively reduce patients' uncertainty. Moreover, this information may undermine patients' trust and have a counterproductive effect [14]. Therefore, it is important to understand how patients' web-based information consumption reduces their uncertainty, so that information providers can improve the design of information to better help patients. With this as the objective, this study intends to answer the following research questions: how do patients' web-based

information-seeking behaviors reduce their uncertainties about diseases and physicians? In addition, how does web-based information, such as information related to diseases and physicians, alleviate problems in this principal-agent relationship and then reduce patients' uncertainty?

To address these research questions, based on the framework of *uncertainty mitigator-uncertainty antecedent-uncertainty*, this study explores how web-based health information mitigates patients' uncertainty. The contributions of this study are as follows. First, based on the principal-agent theory and the uncertainty reduction theory (URT), this study explores the mechanism of how patients' web-based health behaviors can reduce their uncertainty. Second, following the classification of consumers' uncertainty about products and sellers by Dimoka et al [15], this study also distinguishes between patients' uncertainty about diseases and physicians; the influence chain is also investigated. Finally, this study emphasizes the significant role of trust. Additional information can help reduce patients' uncertainty, but only if it can enhance patients' trust in their physicians.

Principal-Agent Theory

Originating from the field of enterprise management, the principal-agent theory describes the relationship in which one entity (the principal) delegates work to another entity (the agent) who performs the work under a mutually agreed contract [16]. The relationship between enterprise owners and professional managers is a typical principal-agent relationship. This relationship applies to all transactional relationships in socioeconomic systems where opportunism, information asymmetry, and limited rationality exist. Owing to the separation of ownership and management rights of enterprises, the goals of principals and agents are inconsistent, which will lead to adverse selection before the contract [17] and the moral hazard of hidden behaviors after the contract [18].

The physician-patient relationship is also a typical principal-agent relationship in which the physician acts as an agent to provide medical services to the patient (the client) under a contract [19,20]. Patients, as principals, receive diagnoses of the disease, treatment plans, and medical care services from the agents (ie, physicians). Physicians and patients have inconsistent goals and asymmetrical information. Compared with physicians, patients are always at a disadvantage in information about diseases and physicians' medical services. Patients want to receive superior medical services at a low cost to improve their health, whereas physicians want to provide medical services at a higher fee and lower cost (to themselves) to increase their income and reputation.

Perceived Information Scarcity

Owing to the principal-agent relationship and the specialization of medical services, there is natural information asymmetry between physicians and patients [21]. Compared with physicians, patients have limited information about diseases and physicians, leading to patients' perception of information scarcity. Previous literature defined scarcity as the limitation or unavailability of objects (eg, commodity) [22]. In the research of Wells et al [23], an individual's degree of prepurchase information scarcity related to the product of interest is operationalized as whether a consumer had any prior information or experience with products offered on web-based shopping websites. Compared with physicians, patients lack professional medical education process and clinical experience; therefore, patients will be aware of the information scarcity regarding diseases and the physician's medical service. In this study, patients' perceived scarcity of information about diseases is defined as patients' perception of their limited information related to diseases, whereas perceived scarcity of information about the physician's medical service information is defined as patients' perception of their limited information related to the physician's medical service.

In the web-based environment, the emergence of information systems can help alleviate the principal-agent problem to some extent. For example, the website and product information can reduce customers' information scarcity about products, thereby reducing customers' worries about the platform's opportunism and their purchase uncertainty [24]; the implementation of information systems within organizations, such as hospitals, was also found to be an effective means of improving information transparency [25]. Similarly, the disease-related and physician-related information obtained by patients through web-based searches can respectively help patients understand diseases and their physicians better. Web-based disease-related and physician-related information can reduce patients' perceived scarcity of information about diseases and their physicians.

However, the information quality is unevenly distributed in the problem of information asymmetry [24], but existing studies failed to take into account the impact of the information quality of search behavior, especially because web-based health information lacks accuracy and credibility [26]. Information quality is always measured by the perceived information quality, which represents information receivers' subjective perception about four dimensions of information quality, namely, relevance, adequacy, usefulness, and understandability of the information [27]. Higher-quality information can lead to better descriptions about the targets, and it is more useful than lower-quality information [28]. With a higher quality of diseases information in the web-based environment, patients will perceive the information as more relevant, adequate, and useful, thereby increasing their information about the diseases. As a result, the higher the quality of disease-related information sought by patients, the lower the perceived scarcity of information about the disease, leading to the following hypothesis: web-based health information quality reduces patients' perceived scarcity of information regarding diseases (H1).

In addition to disease-related information, web-based health information provides patients with physician-related information, such as web-based word-of-mouth information about physicians, which represents other patients' visiting experiences. In traditional offline hospitals, patients had very limited access to physicians' medical service information, which was often confined to the small reach of word-of-mouth communication, making it difficult to obtain a large amount of word-of-mouth physician information. Web-based word-of-mouth information can effectively reduce asymmetries of products information [29,30]. Web-based word-of-mouth information can inform later customers about the details of the products or the service [28,31]. Similarly, physicians' web-based ratings are also found to reflect their quality perceived by offline patients [32]. Web-based word-of-mouth information about physicians obtained by patients before their visit helps patients to know the physicians better, such as the physicians' manner, treatments, and knowledge. Therefore, web-based word-of-mouth information can reduce patients' perceived scarcity of information about their physician's medical services, leading to our second hypothesis: patients' perceived web-based word-of-mouth information about physicians reduces patients' perceived scarcity of information regarding the physicians' medical services (H2).

Fear of Physicians' Opportunistic Behaviors

In the principal-agent relationship, both parties expect to maximize their own interests [16]. The agents will work to increase their benefits, but some of their behaviors may even increase principals' costs, leading to agents' opportunistic behaviors [33]. As principals, patients are concerned about whether the physicians have opportunistic behaviors because patients cannot accurately evaluate physicians' behaviors, especially in China. Owing to the imperfections of the medical systems in China, opportunistic behaviors of medical service providers have caused widespread concerns [34,35], such as whether physicians receive kickbacks, prescribe high-priced drugs [36], or ask patients to do excessive or unnecessary examinations or treatments [36], all of which are beneficial to physicians' own interests but harm patients' interests [37]. Opportunistic behaviors are also harmful to the physician-patient relationship because these behaviors reduce patients' trust in physicians [38].

Patients can not only obtain health information such as diagnoses and treatments through eHealth data but also browse web-based reviews about physicians. Compared with offline word-of-mouth information, web-based word-of-mouth information has a greater impact on consumers' behaviors because of its extensive sources, large coverage, and convenient dissemination [39]. Positive web-based word-of-mouth information can effectively reduce principals' concerns about agents' opportunistic behaviors [24,40]. Web-based word-of-mouth information about physicians helps improve the transparency of medical services and enhance patients' confidence in medical decisions [41]. It also reflects the experiences of other patients with similar diseases [42]. With more web-based word-of-mouth information about the physicians, patients can evaluate the likelihood of the physicians' opportunistic behaviors, and then they can choose physicians who are less likely to engage in those opportunistic

behaviors. Therefore, positive web-based word-of-mouth information helps reduce patients' concerns about physicians' opportunism. Physicians' opportunism, in this study, is defined as the behaviors of physicians who do not provide good services but charge high prices, conduct excessive and unnecessary examinations, and receive rebates to prescribe high-priced drugs [36]. With better web-based word-of-mouth information about physicians, patients will be less apprehensive of the physicians' opportunistic behaviors, leading to the following hypothesis: patients' perceived web-based word-of-mouth information about a physician reduces patients' fear of the physician's opportunism (H3).

Trust

In the principal-agent relationship, trust is the most valuable aspect [43], because if the relationship occurs under ideal conditions, there is no need for trust [44,45]. Trust is the expectation that an individual or a group will make an effort of good faith to behave following commitments (both explicit and implicit), to be honest, and not to take excessive advantage of others, even when the opportunity exists [46]. Owing to the scarcity of patients' information about clinical diagnoses and treatments, the asymmetry of physicians' medical service information between patients and physicians makes it difficult for patients to determine whether the physicians are trustworthy [24]; therefore, in the physician-patient relationship, the information scarcity of physicians' medical services impedes patients' trust in the physicians, leading to hypothesis 4: patients' perceived information scarcity about physicians' medical service information reduces patients' trust in physicians (H4).

In the principal-agent relationship, as agents, patients' fear of physicians' opportunistic behaviors also influences patients' trust in physicians. Existing research has confirmed that opportunistic behavior in web-based banking leads to low levels of trust of users in internet banking [47]. In the e-commerce environment, fear of sellers' opportunism also harms buyers' trust [33]. In the physician-patient relationship, opportunistic behaviors are also harmful because these behaviors reduce patients' trust in physicians [38]. Although physicians' behaviors are not always immoral, patients still worry about the possibility of physicians' opportunistic behaviors because the principal-agent relationship is favorable for physicians to act immoral behaviors. This worry will be enhanced if the possibility of the physicians' opportunism is high. Physicians' opportunistic behavior benefits their own interests but harms the interests of patients, which also impedes patients' trust in them. Patients cannot monitor physicians' behaviors, and they worry that their physicians will act opportunistic behaviors; thus, the fear of physicians' opportunism reduces patients' trust, leading to hypothesis 5: the fear of physicians' opportunism reduces patients' trust in physicians (H5).

Web-based word-of-mouth information is an important factor affecting potential customers' purchase intentions and behaviors [48,49], because web-based word-of-mouth information reflects previous consumers' evaluation of the products. In medical situations, some studies have also explored the impact of web-based physician reviews on patients' decision-making

behavior. For example, higher web-based ratings of physicians increase patients' intention to consult them [50]. Web-based word-of-mouth information about physicians also increases physicians' offline visits [51]. Acting as the previous patients' evaluation cue, physicians' web-based word-of-mouth information serves as an important reference for the selection of physicians by patients. The better the patients perceive web-based word-of-mouth information about the physicians, the more favorable it is for the patients to trust in the physicians, leading to the following hypothesis: patients' perceived web-based word-of-mouth information about physicians increases patients' trust in the physicians (H6).

URT Overview

In the principal-agent relationship, uncertainty arises because the principal cannot fully monitor the agent's behavior, resulting in adverse selection [17] and the moral hazard of hidden behaviors [18]. It is important to understand how to reduce uncertainty in this relationship. For example, reducing uncertainty can increase consumers' purchase intention and lead to an actual purchase [24]; reducing uncertainty can also increase users' trust in the web-based world so that they can effectively use a tool [52]. Originating from the field of interpersonal communication, the URT posits that uncertainty occurs when people cannot predict the future behavior of others or when they do not meet their own expectations [52,53]. URT is widely used in fields such as organizational behavior and information systems, among others [52]. For example, Srivastava and Chandra [52] considered 3 ways to reduce users' uncertainty to enhance their trust and use intention in the web-based world. The three ways include acquiring information passively through observation, acquiring information actively through third-party search, digital signatures, and third-party authentication, and acquiring information from interactions, such as direct interaction with the target object [52].

In the medical scenario, patients' uncertainties, that is, their inability to accurately predict the state of their disease because of a lack of information, exist in every aspect of their diagnoses and treatments. Uncertainties in the principal-agent relationship are caused by specific information problems (eg, hiding information and hiding behavior), and these problems can be alleviated by the use of information systems [24]. In the physician-patient relationship discussed previously, the disclosure of information comes from the agent (eg, medical information provided by the physician), and it reduces only a few uncertainties of patients, but with the development of technology, medical and health information is no longer only in the hands of the medical providers (agents). The client can actively acquire medical and health information from a third party [20], enabling patients to overcome the restrictions of time and space and actively obtain information about the causes of diseases, treatments, and reputations of physicians and hospitals through the internet. With the active information acquisition method [52] to reduce uncertainty, patients' web-based search behavior can help actively reduce uncertainty, but the influence mechanism of how web-based information acquired by patients reduces uncertainty is not yet clear.

Patients' Uncertainty

Uncertainty in the medical context refers to a cognitive state in which the meaning of medical events cannot be determined [3,4]. Uncertainty, as a medical experience characterized by unpredictability, unfamiliarity, and ambiguity, is associated with poor medical outcomes and psychological states (eg, fear, stress, and loss of control) [43]. Existing research on uncertainties in the medical field has mainly focused on information uncertainty related to diseases, diagnoses, and treatments [5]. Uncertainties regarding illness can be divided into the medical providers' uncertainty about diseases and the patients' uncertainty about diseases. Previous research has mainly focused on the physician's uncertainty of expressing disease-related information during patients' visits and its impact [54]. The latter, that is, patients' uncertainty about diseases, is the focus of this study.

Patients' uncertainty means that the patients are unable to determine the meaning of disease-related events or accurately predict the outcomes of such events [5,6]. In the uncertainty in illness theory presented by Mishel [3], the antecedents (eg, symptom stimulus, patients' cognitive abilities, and physicians' information authorities), the appraisal process, the coping mechanism, and the adaptation outcomes of patients' uncertainty in diseases are concluded, and the scale of patient uncertainty about illness is developed. This theory is effective in guiding interventions to manage patients' uncertainty [55].

In web-based markets, as sellers cannot fully describe the product or predict the products' future performances, consumers' uncertainty about products and sellers should be distinguished, between which the former uncertainty is related to the description and performance of products, and the latter uncertainty is related to sellers' adverse selection and moral hazard [15]. The uncertainty about sellers also increases uncertainty about products, and the two types of uncertainties reduce price premiums [15]. Similarly, in the physician-patient relationship, as physicians cannot fully describe the diseases or predict the effectiveness of treatments, patients' uncertainty in the process may be not only about the diseases but also about the physicians. Owing to the traits of principal-agent relationship, patients, who are the inferior party because of the scarcity of information, tend to question the rationality of physicians' advised medical treatments. However, in the medical context, few researchers have focused on patients' uncertainty about physicians. Given this, considering the principal-agent relationship between physicians and patients, we follow the classification of customers' uncertainties about sellers and products by Dimoka et al [15] to distinguish between patients' uncertainty about diseases and patients' uncertainty about physicians. In this way, this study can contribute to research on patients' uncertainty.

In the principal-agent relationship, how much information principals have played a key role in their uncertainty [23,24]. The lower the availability of product information, the greater the consumers' uncertainty about the product quality [23]; therefore, in our context, patients' perception of scarcity of disease information can increase patients' uncertainties about the diseases, and we hypothesize the following: patients'

perceived information scarcity about diseases increases patients' uncertainty about the diseases (H7).

Owing to information scarcity, it is difficult for patients to judge the quality of physicians' medical services. Less information about physicians' medical services leads to patients' stronger sense of uncertainty about physicians. According to research on the uncertainty of patients regarding disease [3], the causative factors include event familiarity. When patients have more knowledge about the physicians' medical services, it helps to reduce their uncertainty about the physicians' medical services, leading to the following hypothesis: patients' perceived information scarcity about physicians' medical services increases patients' uncertainty about the physicians (H8).

Trust can overcome uncertainty, and trust is necessary only when the environment is uncertain [45]. When patients trust their physicians, they can predict their physicians' behaviors based on their belief in the physicians' integrity, benevolence, and competence under uncertain circumstances. They believe that their physicians are honest and have great capabilities. Therefore, this study believes that a patient's trust in a physician will help reduce the patient's uncertainty about the physician, leading to hypothesis 9: a patient's trust in a physician can mitigate the patient's uncertainty in that physician (H9).

Because of the internally inconsistent goals between physicians and patients, physicians' opportunistic behaviors are inevitable, such as physicians taking kickbacks to prescribe expensive drugs, unnecessary tests, and overtreatment. Patients often lack professional information to judge the rationality of physicians' treatment plans and examination procedures, which leads to a sense of uncertainty about the rationality of physicians' treatment behaviors. In China, concern about physicians' opportunistic behavior is an important factor that leads to patients' sense of uncertainty [37,56]. Possible opportunistic behavior of vendors' drug prescription also leads to more uncertainty for buyers [45], and thus we hypothesize the following: patients' fear of the physician's opportunism increases patients' uncertainty in the physician (H10).

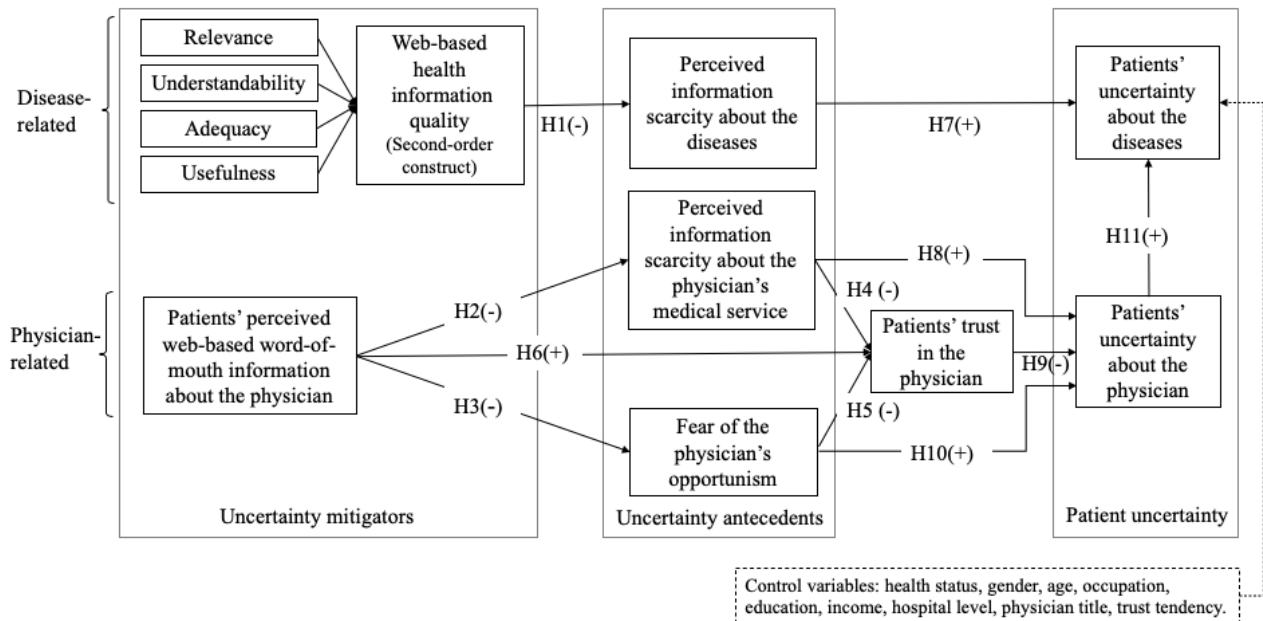
Consumers' uncertainty about sellers is distinct from the uncertainty about products, between which the former uncertainty can increase the latter uncertainty [15]. The process of patient consultation entails providing a series of examinations, diagnoses, and other services by the physician to identify the disease and determine other relevant treatments for the patient. If patients are uncertain about the rationality of the medical services provided by physicians and doubt the rationality of the physician's examination and treatment plans, it will be detrimental to patients' certainty about the disease; therefore, we hypothesize the following: patients' uncertainty about the physician increases patients' uncertainty about their diseases (H11).

In summary, based on existing literature, this study uses the principal-agent theory and the URT to develop a research model to explain the mechanism of how web-based information search by patients can reduce their uncertainties, as shown in Figure 1. In the context of patients' active information acquisition, we hypothesize that two types of web-based health information (ie,

web-based disease-related information quality and web-based physician-related information) as uncertainty mitigators to reduce patients' uncertainty. When identifying the antecedents of uncertainty and its consequences, we followed the descriptions by Pavlou et al [24] and Srivastava and Chandra [52] on the use of unique and specific variables related to customers' uncertainty. Principals' perception of information scarcity and principals' concern about agents' opportunistic

behaviors are the causes of the principals' uncertainty. Principals' trust in the agents acts as a mediator when the uncertainty antecedents reduce principals' uncertainty. Following the classification of customers' uncertainty about sellers and products by Dimoka et al [15], we classified patients' uncertainties into uncertainty about diseases and uncertainty about physicians. Our research model is depicted in Figure 1.

Figure 1. Research conceptual model. H: hypothesis.



Methods

Ethical Considerations

An ethics review was not applicable for this study because the online survey measured the subjects' perceptions and did not influence their perceptions or attitudes.

Data Collection

This study adopted the survey method to collect data. A total of 108 questionnaires were collected for the pilot test before the formal survey. The wording of some items and typesetting in the questionnaire were modified according to the feedback of the participants. A professional survey company (Wenjuanxing) was responsible for collecting the formal data. The survey started in May 2020 and lasted for a month. Each questionnaire corresponding to a separate IP address provided a reward of RMB 14 (US \$2.20). The questionnaire was also set to ensure that valid respondents should answer all the questions before submitting. At the beginning of the questionnaire, the background of the survey was introduced, and screening questions were set to meet the 3 requirements for the survey. Only those who might have a certain disease and have seen a physician offline within 3 months, who had engaged in web-based disease information search behavior, and who had read the web-based word-of-mouth information of the visited physicians were eligible.

Specifically, our questionnaire first used 3 questions to exclude invalid respondents. The first question was "Did you suffer a

certain disease and have any experience of offline medical treatments in the past three months?" The respondents who answered "Yes" proceeded to the next question, and the respondents who answered "No" were regarded as invalid respondents, and their questionnaires were terminated. Then, at the top of each page, there was a statement "Please recall the most recent experience of seeing a physician within the past three months, and based on this experience, answer the following questions." The second question to screen out the invalid respondents was "Before the consultation, have you searched for disease-related information on the internet for this consultation?" Similarly, respondents who answered "No" were prompted to end answering the questionnaire, and those who answered "Yes" continued to the next question. The third screening item was "Do you know the evaluation of the physician (based on the most recent visit within three months) on the Internet?" Only respondents who answered "Yes" continued to answer the questions about the perception of web-based word-of-mouth information, and respondents who answered "No" were prompted to end answering the questionnaire.

To ensure that the respondents responded seriously, the question regarding the evaluation of physicians' medical services appeared twice in different places in the questionnaire. Questionnaires with completely inconsistent answers (eg, strongly disagree and very agree) were excluded. A total of 40 invalid respondents were screened out, and the final sample size was 337. This sample size meets the requirement that the sample

size should be 5-10 observations for each estimated parameter [57,58].

Measurements

All items in this study are from mature scales, as shown in Table 1. Web-based information quality is a formative construct and the measurements were from Zahedi and Song [59]. The modified scale for perceived web-based word-of-mouth information about physicians was derived from Collins and Stevens [60]. As mentioned previously, there is a filter item—"Do you know the evaluation of the physician (based on the most recent visit within three months) on the Internet?"—which asked the respondent whether he or she had browsed through the web-based word-of-mouth information about the physician from those who had previously consulted that physician. With this filter item, we could ensure that the respondent's answers to web-based word-of-mouth information and other items were for the same physician. The measurement of perceived information scarcity about disease and physicians' medical service was derived from Wells et al [23], who developed reflective measures to assess individuals' degree of prepurchase information scarcity about products. The measurement of fear of physicians' opportunism was from the measurement of fear of sellers' opportunism [24,40], which referred to patients' concerns about the rationality of the visited

physicians' treatment behaviors (eg, excessive examination and high-priced drugs). Patients' trust measurement was modified from that suggested by McKnight et al [61] and Zhou et al [62]. Patients' uncertainty about diseases was measured using the community scale of uncertainty in illness (Mishel Uncertainty in Illness Scale–Community form) [3]. Patients' uncertainty about physicians was modified from the perceived uncertainty scale [63], which referred to patients' uncertainty about the rationality of medical services provided by physicians. Respondents in this study are native Chinese speakers; therefore, all items were translated into Chinese. We conducted translation–back-translation procedure to ensure the validity of our questionnaire. Specifically, the translated questionnaire was evaluated by 2 doctoral students with relevant research backgrounds. Some adjustments were made to the wording and expression of the questionnaire based on their feedback. Items of constructs (ie, perceived web-based word-of-mouth information about physicians, perceived information scarcity, fear of the physician's opportunism, perceived uncertainty, and trust) were measured by a 5-point Likert scale ranging from complete disagreement (1) to complete agreement (5). Items of the 4 dimensions of information quality were measured by the extent to which the internet health information conforms to the description in the item (eg, 1 point for a very low level and 5 points for a very high level).

Table 1. Construct measurement.

Construct, label, and source	Item
IQ^a [59]	
Relevance1	For your health information needs, to what degree do you believe the internet health information provided by the website was applicable to your needs?
Relevance2	For your health information needs, to what degree do you believe internet health information provided by the website was related to your needs?
Relevance3	For your health information needs, to what degree do you believe internet health information provided by the website was pertinent to your needs?
Relevance4	For your health information needs, to what degree do you believe internet health information provided by the website was relevant to your needs?
Understandability1	For your health information needs, to what degree do you believe internet health information provided by the website was clear in meaning?
Understandability2	For your health information needs, to what degree do you believe internet health information provided by the website was easy to read?
Understandability3	For your health information needs, to what degree do you believe internet health information provided by the website was easy to comprehend?
Understandability4	For your health information needs, to what degree do you believe internet health information provided by the website was understandable?
Adequacy1	For your health information needs, to what degree do you believe internet health information provided by the website was sufficient?
Adequacy2	For your health information needs, to what degree do you believe internet health information provided by the website was complete?
Adequacy3	For your health information needs, to what degree do you believe internet health information provided by the website was adequate?
Adequacy4	For your health information needs, to what degree do you believe internet health information provided by the website contained the necessary topics or categories?
Usefulness1	For your health information needs, to what degree do you believe internet health information provided by the website was informative?
Usefulness2	For your health information needs, to what degree do you believe internet health information provided by the website was valuable?
Usefulness3	For your health information needs, to what degree do you believe internet health information provided by the website was helpful?
Usefulness4	For your health information needs, to what degree do you believe internet health information provided by the website was useful?
PWOM^b [60]	
PWOM1	In online reviews, the physician is very popular and many patients come to see the physician.
PWOM2	In online reviews, patients who visited the physician had a good experience.
PWOM3	According to online reviews, the physician is a good physician.
PWOM4	According to online reviews, the physician has a good relationship with patients.
PSD^c [23]	
PSD1	I have a good idea of the disease-related information (eg, symptoms, causes of disease, treatment methods, etc).
PSD2	I have sufficient information about the disease (eg, symptoms, cause of disease, treatment, etc).
PSD3	I possess adequate knowledge about the disease-related information (eg, symptoms, causes of disease, treatment methods, etc).
PSPMS^d [23]	
PSPMS1	I have a good idea of the medical services of the physician whom I visited this time.
PSPMS2	I have sufficient information about the medical services of the physician for this visit.
PSPMS3	I possess adequate knowledge about the medical service information of the physician whom I visited this time.

Construct, label, and source	Item
FPO^e [24]	
FPO1	In this visit, the physician might not have provided good service but charged a high price.
FPO2	In this visit, the physician might have overexamined, unnecessarily examined, or overtreated me.
FPO3	In this visit, the physician might have received a rebate for prescribing an overpriced drug (eg, imported drug).
FPO4	In this visit, the physician might have breached formal or informal agreements to his or her benefit.
T^f [61,62]	
T1	The physician is sincerely concerned about my medical issues
T2	The physician is honest in his or her medical practices
T3	I believe that the physician does a very good job
T4	I feel that I can count on the physician to help me with my medical problems
MUIS^g [3]	
MUIS1	I don't know what is wrong with me
MUIS2	I have a lot of questions without answers.
MUIS3	It is difficult to know if the treatments or medications I am getting are helping.
MUIS4	Because of the unpredictability of my illness, I cannot plan for the future.
MUIS5	The effectiveness of the treatment is undetermined.
PU^h [63]	
PU1	I think the rationality of the medical services provided by the physician involves a high degree of uncertainty.
PU2	I think the rationality of the medicine prescribed by the physician is uncertain.
PU3	I think the rationality of the disease examination and treatment plan is uncertain.
PU4	The rationality of the services provided by the physician is uncertain (ie, the service I received may not be exactly what I wanted).
PU5	I feel the uncertainty associated with the rationality of the medical services provided by the physician is high.
TDⁱ [64]	
TD1	I generally trust other people.
TD2	I generally have faith in humanity.
TD3	I feel that people are generally reliable.
TD4	I generally trust other people unless they give me reasons not to.

^aIQ: web-based health information quality.

^bPWOM: perceived web-based word-of-mouth information about physicians.

^cPSD: perceived information scarcity about the diseases.

^dPSPMS: perceived information scarcity about the physicians' medical services.

^eFPO: fears of physician's opportunism.

^fT: patients' trust in the physician.

^gMUIS: patients' uncertainty about diseases.

^hPU: patients' uncertainty about the physician.

ⁱTD: trust tendency.

To reduce other possible influences on our model, we considered control variables in 3 ways, although these variables are not our interest in this study. To reduce the possible influence of individual differences, demographic information, such as gender, age, education level, income per month, and occupation, is controlled. To reduce the possible influence of the impact of medical treatment, health-related and medical experience-related factors are also controlled, such as the respondent's health status,

the physician's title (an official certification of a physician's quality by the government) [65], and the hospital's level (an official certification of a hospital's quality by the government) [65]. To reduce the possible influence of the respondent's characteristic of trust, the respondents' trust tendency was also controlled. For example, with the same word-of-mouth information about a physician, some patients may easily trust the physician, whereas others may still doubt the physician.

Trust tendency [64] was also measured by a 5-point Likert scale ranging from complete disagreement (1) to complete agreement (5).

Results

Overview

As the model measured in this study has a formative construct, partial least squares (PLS) structural equation modeling is suitable for data analysis. SmartPLS (version 3.3.3, SmartPLS GmbH) software was used in this study. In addition, PLS is also widely used in information systems research owing to its relaxed requirements for the normal distribution of samples, its ability to process data with small sample size, and its applicability to development theory rather than test theory [66]. We first used SmartPLS (version 3.3.3, SmartPLS GmbH) software to test

the reliability and validity of the measurement model and then tested the path coefficients of the structural model.

Descriptive Statistics

The respondents' demographic information, health-related information, and medical experience-related information are shown in Table 2. More respondents were female (231/337, 68.5%). In terms of age distribution, age groups 21-30 years (165/337, 49%) and 31-40 years (117/337, 34.7%) were the most represented. Education level was relatively high, with high school and below accounting for only 4.5% (15/337). The monthly income distribution was relatively even. The surveyed samples were mainly working people, with enterprise employees accounting for 68.2% (230/337). The physicians' titles and hospitals' levels are also relatively high.

Table 3 lists the descriptive statistics of the constructs involved in the model.

Table 2. Demographic profile, health-related information, and medical experience–related information (N=337).

Characteristic	Value, n (%)
Gender	
Female	231 (68.5)
Male	106 (31.5)
Age (years)	
18-20	23 (6.8)
21-30	165 (49)
31-40	117 (34.7)
41-50	32 (9.5)
Education	
Postgraduate or above	25 (7.4)
Undergraduate	246 (73)
3-year college	51 (15.1)
High school	11 (3.3)
Middle school or below	4 (1.2)
Monthly income (RMB [US \$])	
≤3000 (471.60)	55 (16.3)
3000-5999 (471.60-943.20)	40 (11.9)
6000-8999 (943.20-1414.80)	87 (25.8)
9000-11,999 (1414.80-1886.40)	85 (25.2)
12,000-14,999 (1886.40-2358)	43 (12.8)
≥15,000 (2358)	27 (8)
Occupation	
Student	46 (13.6)
Enterprise worker	230 (68.2)
Civil servant	39 (11.6)
Individual operator	15 (4.5)
Others	7 (2.1)
Health status	
Excellent	14 (4.1)
Very good	56 (16.6)
Good	124 (36.8)
Fair	134 (39.8)
Poor	9 (2.7)
Physician's title	
Assistant physician	103 (30.6)
Associate physician	123 (36.5)
Chief physician	87 (25.8)
Not sure	24 (7.1)
Hospital's level	
Primary hospital	52 (15.4)
Intermediate hospital	72 (21.4)
Senior hospital	203 (60.2)

Characteristic	Value, n (%)
Not sure	10 (3)

Table 3. Descriptive statistics^a.

Construct and item	Minimum value	Maximum value	Mean (SD)
PWOM^b			
PWOM1	2	5	4.04 (0.66)
PWOM2	1	5	4.16 (0.80)
PWOM3	1	5	4.08 (0.88)
PWOM4	1	5	4.09 (0.85)
PSD^c			
PSD1	1	5	2.38 (0.83)
PSD2	1	5	2.62 (1.03)
PSD3	1	5	2.48 (0.97)
PSPMS^d			
PSPMS1	1	5	2.24 (0.76)
PSPMS2	1	5	2.25 (0.90)
PSPMS3	1	5	2.23 (0.87)
FPO^e			
FPO1	1	5	2.44 (1.04)
FPO2	1	5	2.40 (1.19)
FPO3	1	5	2.06 (1.07)
FPO4	1	5	1.90 (1.08)
T^f			
T1	1	5	4.03 (0.76)
T2	2	5	4.12 (0.78)
T3	1	5	4.02 (0.79)
T4	1	5	4.10 (0.81)
MUIS^g			
MUIS1	1	5	2.27 (0.94)
MUIS2	1	5	2.70 (1.12)
MUIS3	1	5	2.60 (1.15)
MUIS4	1	5	2.35 (1.14)
MUIS5	1	5	2.64 (1.05)
PU^h			
PU1	1	5	2.63 (1.04)
PU2	1	5	2.28 (1.13)
PU3	1	5	2.32 (1.18)
PU4	1	5	2.43 (1.13)
PU5	1	5	2.34 (1.08)
TDⁱ			
TD1	1	5	3.73 (0.74)
TD2	1	5	3.91 (0.78)
TD3	1	5	3.72 (0.90)
TD4	1	5	3.78 (0.91)

^aThe web-based health information quality is a formative construct; therefore, the details of this construct are described in the *Measurement Model* section.

^bPWOM: perceived web-based word-of-mouth information about physicians.

^cPSD: perceived information scarcity about the diseases.

^dPSPMS: perceived information scarcity about the physicians' medical services.

^eFPO: fears of physician's opportunism.

^fT: patients' trust in the physician.

^gMUIS: patients' uncertainty about diseases.

^hPU: patients' uncertainty about physicians.

ⁱTD: trust tendency.

Common Method Variance

As with all self-reported data, we should examine the potential common method variance. We follow the suggestions of Podsakoff et al [67] to minimize potential common method biases. First, we tried procedural remedies of Podsakoff et al [67]. To reduce respondents' evaluation apprehension and avoid their answers being socially desirable, at the beginning of the questionnaire, we reminded them that their answers are anonymous and there are no right or wrong answers to our questions. All items in the questionnaire were designed in a random order to ensure that the measurement of predictor and criterion variables are psychologically separated for respondents. To ensure that the scale items are specific, concise, and clear, we also conducted the pilot test before the formal survey. We modify the wording according to the feedback of the participants to reduce ambiguity.

Second, the Harman single-factor test was conducted to diagnose whether the common method bias is a problem [68]. We ran an exploratory factor with all variables included [23]. The results showed that more than one factor can be extracted from the unrotated solution, and the variance contribution rate of the first factor was not more than 50% (23.7%), so there was no one single major factor that can reflect the majority covariance of all items, indicating that common method bias was not serious [57].

Moreover, based on our survey context and the suggestions of Podsakoff et al [67], we conducted a single-common-method-factor approach by controlling for the effects of a single unmeasured latent method factor to control the common method variance. Following Liang et al [69], we included in the PLS model a common method factor whose indicators included all the indicators of the constructs in this study. We calculated each indicator's factor loadings and variances substantively explained by the construct and by the method factor. [Multimedia Appendix 1](#) provides the detailed procedure and results [67,69,70]. As shown in Table S1 in [Multimedia Appendix 1](#), most factor loadings of the method factor are insignificant. The average substantively explained variance of the indicators is 0.594, whereas the average method-based variance is 0.002. The ratio of substantive variance to method variance was 297:1, indicating the variance of each observed indicator explained by its substantive construct

is substantially greater than the variance explained by the method factor. Therefore, based on the studies by Liang et al [69] and Williams et al [70], we further conclude that common method bias is not a serious problem in this study.

Measurement Model

First, we tested the reliability and validity of the formative indicators (ie, web-based information quality). As web-based health information quality is a second-order formative construct, this study follows the method suggested by Wetzels et al [71]. In the structural equation model, four first-order reflective constructs (ie, information relevance, understandability, adequacy, and usefulness) point to the second-order constructive variable (information quality). A total of 16 items in the first order are taken as the measurement items of second-order constructs. PLS and Bootstrap were used to test the reliability and validity of the model and the outer weight of second-order formative constructs. First, the results of reliability and validity test of first-order reflective constructs showed information relevance (Cronbach α =.641; composite reliability [CR]=0.786; average variance extracted [AVE]=0.480), information understandability (Cronbach α =.726; CR=0.830; AVE=0.551), information usefulness (Cronbach α =.699; CR=0.816; AVE=0.526), and information adequacy (Cronbach α =.868; CR=0.910; AVE=0.717) all have good reliability and validity. Then, we tested the reliability and validity of the information quality of the second-order formative index, and the weight of the information quality (0.263, 0.314, 0.293, and 0.463) was >0.2 and significant at the level of $P<.001$, which passed the reliability and validity test of the formative construct. The variance inflation factors among all items were <2, satisfying the multicollinearity test, and the outer weight was significant and >0.2 [72].

Second, reflective indicators of this model were tested. We followed the methods suggested by Lewis et al [58] and Straub et al [73] to test the reliability and validity of the measurement model. The results are listed in [Table 4](#). First of all, we tested the reliability of the constructs. The results show that the component reliability of each construct is >0.7 with good internal consistency [74,75]. The average variance extraction is also >0.5, which has good convergent validity [76]. In most cases, Cronbach α is >.7, and in all cases, the values are >0.6, which are within the acceptable range [66].

Table 4. Construct reliability and validity.

Construct and item	Item loading	Cronbach α	CR ^a	AVE ^b
PWOM^c		.653	0.793	0.500
PWOM1	0.703			
PWOM2	0.695			
PWOM3	0.712			
PWOM4	0.689			
PSD^d		.740	0.852	0.658
PSD1	0.797			
PSD2	0.832			
PSD3	0.805			
PSPMS^e		.695	0.830	0.620
PSPMS1	0.758			
PSPMS2	0.784			
PSPMS3	0.820			
FPO^f		.852	0.900	0.692
FPO1	0.821			
FPO2	0.811			
FPO3	0.863			
FPO4	0.831			
PU^g		.890	0.919	0.694
PU1	0.861			
PU2	0.828			
PU3	0.828			
PU4	0.803			
PU5	0.844			
MUIS^h		.797	0.861	0.554
MUIS1	0.663			
MUIS2	0.772			
MUIS3	0.790			
MUIS4	0.706			
MUIS5	0.783			
Tⁱ		.731	0.832	0.554
T1	0.710			
T2	0.759			
T3	0.805			
T4	0.699			
TD^j		.760	0.846	0.580
TD1	0.834			
TD2	0.796			
TD3	0.740			
TD4	0.666			

^aCR: composite reliability.

^bAVE: average variance extracted.

^cPWOM: perceived web-based word-of-mouth information about physicians.

^dPSD: perceived information scarcity about the diseases.

^ePSPMS: perceived information scarcity about the physicians' medical services.

^fFPO: fears of physician's opportunism.

^gPU: patients' uncertainty about physicians.

^hMUIS: patients' uncertainty about diseases.

ⁱT: patients' trust.

^jTD: trust tendency.

As shown in Table 5, we also tested the discriminant validity of the measurement model. The square root of the AVE (ie, italicized number on the diagonal line) for each factor in the table is larger than the correlation coefficient between the factor

and other factors, so this measurement model has good discriminant validity [76]. Therefore, all the reflective constructs of this measurement model have good reliability and validity.

Table 5. Discriminant validity analysis^a.

Construct	IQ ^b	PU ^c	PSPMS ^d	PWOM ^e	T ^f	FPO ^g	MUIS ^h	PSD ⁱ
IQ	<i>.—</i> ^j	—	—	—	—	—	—	—
PU	-0.323	<i>0.833</i>	—	—	—	—	—	—
PSPMS	-0.506	0.258	<i>0.788</i>	—	—	—	—	—
PWOM	0.405	-0.336	-0.511	<i>0.700</i>	—	—	—	—
T	0.379	-0.539	-0.473	0.532	<i>0.744</i>	—	—	—
FPO	-0.154	0.711	0.118	-0.268	-0.380	<i>0.832</i>	—	—
MUIS	-0.365	0.678	0.301	-0.279	-0.497	0.507	<i>0.744</i>	—
PSD	-0.588	0.255	0.487	-0.273	-0.334	0.068	0.296	<i>0.811</i>

^aThe italicized values represent the square root of the average variance extracted for each construct.

^bIQ: web-based health information quality.

^cPU: patients' uncertainty about the physician.

^dPSPMS: perceived information scarcity about the physicians' medical services.

^ePWOM: perceived web-based word-of-mouth information about physicians.

^fT: patients' trust in the physician.

^gFPO: fears of physician's opportunism.

^hMUIS: patients' uncertainty about diseases.

ⁱPSD: perceived information scarcity about the diseases.

^jNot applicable.

Construct Model and Results

We used PLS to test the hypotheses of this model and the Bootstrap method to test the significance of path coefficients [77]. The results are shown in Figure 2, and the path coefficients and T values are shown in Table 6. The control variables were also included in the model as predictors of the finally dependent variable (ie, patients' uncertainty about the diseases). From Figure 2, the R² of this model for patients' uncertainty in diseases is 0.515. Both the disease- and physician-related uncertainty mitigators have significant effects on the uncertainty

antecedents. Specifically, web-based health information quality can reduce patients' perceived information scarcity about diseases ($\beta=-.588$; $P<.001$), supporting H1. Patients' perceived web-based word-of-mouth information about physicians can reduce patients' perceived information scarcity about the physician's medical service ($\beta=-.511$; $P<.001$) and fears of physicians' opportunism ($\beta=-.268$; $P<.001$), thus supporting H2 and H3. Patients' perceived web-based word-of-mouth of physicians also increases patients' trust in the visited physician ($\beta=.318$; $P<.001$), supporting H6.

Figure 2. Structural equation model results. H: hypothesis.

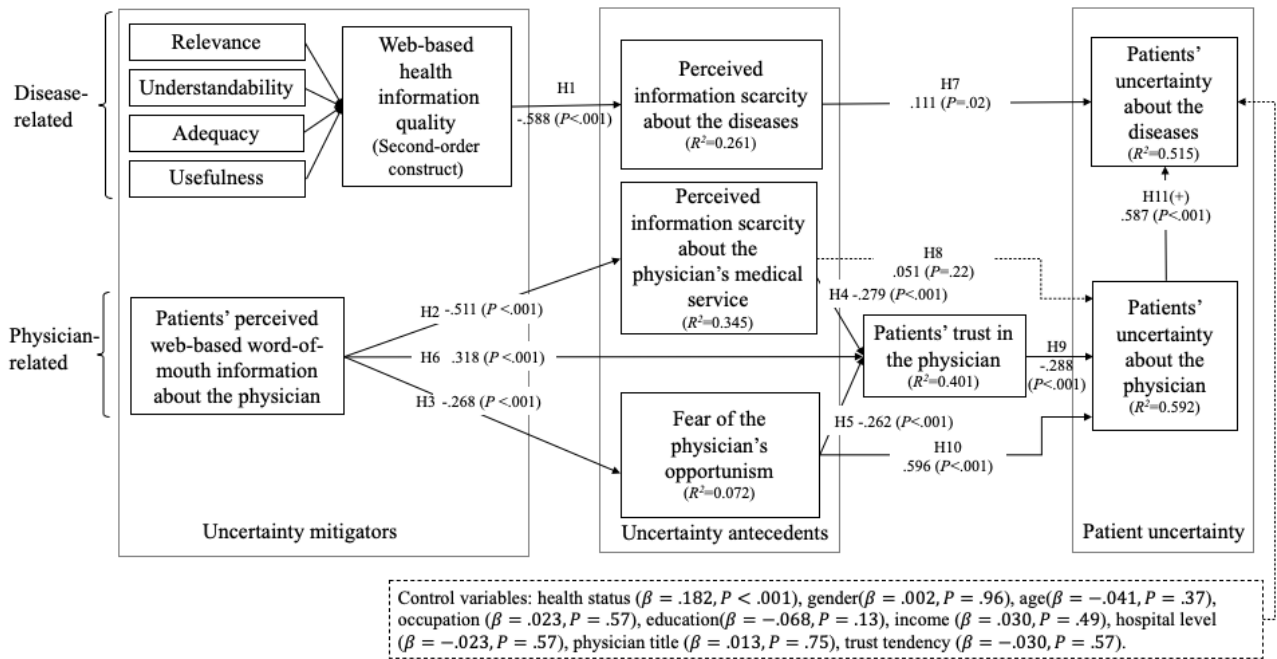


Table 6. Hypotheses test results.

Hypothesis	Path	Path coefficient (SD)	P value	Result
H1	IQ ^a →PSD	-.588 (0.035)	<.001	Supported
H2	PWOM ^b →PSPMS ^c	-.511 (0.045)	<.001	Supported
H3	PWOM→FPO ^d	-.268 (0.048)	<.001	Supported
H4	PSPMS→T ^e	-.279 (0.062)	<.001	Supported
H5	FPO→T	-.262 (0.050)	<.001	Supported
H6	PWOM→T	.318 (0.068)	<.001	Supported
H7	PSD ^f →MUIS ^g	.111 (0.045)	.02	Supported
H8	PSPMS→PU ^h	.051 (0.045)	.22	Rejected but fully mediated by patients' trust in the physician
H9	T→PU	-.288 (0.043)	<.001	Supported
H10	FPO→PU	.596 (0.047)	<.001	Supported and partially mediated by patients' trust in the physician
H11	PU→MUIS	.587 (0.043)	<.001	Supported

^aIQ: web-based health information quality.

^bPWOM: perceived web-based word-of-mouth information of the physician.

^cPSPMS: perceived information scarcity about the physicians' medical services.

^dFPO: fears of physician's opportunism.

^eT: patients' trust in the physician.

^fPSD: perceived information scarcity about the diseases.

^gMUIS: patients' uncertainty about diseases.

^hPU: patients' uncertainty about the physician.

Patients' perceived information scarcity about the physician's medical service reduces their trust in the visited physician ($\beta = -.279$; $P < .001$), supporting H4. Fear of physicians' opportunism reduces patients' trust in the visited physician ($\beta = -.262$; $P < .001$), supporting H5. Patients' perceived information scarcity about the diseases increases patients' uncertainty in diseases ($\beta = .111$; $P = .02$), supporting H7.

However, patients' perceived information scarcity about physicians' medical services has no significant influence on patients' uncertainty in the visited physician ($\beta = .051$; $P = .22$), thus rejecting H8. Patients' trust in the visited physician can reduce patients' uncertainty in the visited physician ($\beta = -.288$; $P < .001$), supporting H9. Fear of physicians' opportunism has the most significant positive effect to increase patients'

uncertainty about the physician ($\beta=.596$; $P<.001$), supporting H10. Finally, uncertainty about the visited physician can increase patients' uncertainty in diseases ($\beta=.587$; $P<.001$), supporting H11.

Besides respondents' perception about their health status, other control variables have no significant influence on the model. Health status has a significantly negative impact on the model, which means that compared with patients who feel their health status is poor, patients who feel they are healthy perceive a higher level of uncertainty about the diseases.

To further explore the possible explanation of the rejection of H8, we conducted the Sobel test [78,79] to investigate the mediation role of trust in the relationship between the uncertainty antecedents and patients' uncertainty about the physician. From the results in Table 7, after introducing patients'

trust in their physicians, the relationship between patients' perceived information scarcity about physicians and their uncertainty about the physicians becomes nonsignificant, indicating that patients' trust in their physicians fully mediates the relationship of H8; therefore, the direct relationship of H8 is rejected, and only when more information can increase patients' trust, their uncertainty about physicians can be reduced. Moreover, increasing physicians' medical service information can be effective in reducing patients' uncertainty about their physicians. The relationship between patients' fear of the physician's opportunism and their uncertainty about the physician is still significant, indicating patients' trust in their physicians partially mediates the relationship of H10. The Sobel test statistics [80] are also significant, which further confirms that patients' trust in their physicians plays the role of mediation.

Table 7. The Sobel test of the mediating effect of patients' trust in the physician.

Hypothesis and path	Path coefficient (SD)	Sobel test statistic	P value
Hypothesis 8^a	N/A ^b	6.5734	<.001
PSPMS^c→PU^d			
Without mediator	.120 (0.054)	N/A	.02
With mediator	-.055 (0.052)	N/A	.29
PSPMS→T ^e	-.476 (0.053)	N/A	<.001
T→PU	-.492 (0.051)	N/A	<.001
Hypothesis 10^f	N/A	5.2280	<.001
FPO^g→PU			
Without mediator	.656 (0.040)	N/A	<.001
With mediator	.579 (0.047)	N/A	<.001
FPO→T	-.380 (0.047)	N/A	<.001
T→PU	-.281 (0.041)	N/A	<.001

^aFully mediated.

^bN/A: not applicable.

^cPSPMS: perceived information scarcity about the physicians' medical services.

^dPU: patients' uncertainty about the physician.

^eT: patients' trust in the physician.

^fPartially mediated.

^gFPO: fears of physician's opportunism.

Discussion

Principal Findings

This study investigated the mechanism of how web-based health information search behavior reduces patients' uncertainty. Our empirical test results supported most of our hypotheses, except H8. Patients' perceived web-based word-of-mouth information about physicians and the quality of web-based health information can effectively reduce patients' uncertainty about diseases and physicians. The uncertainty reduction effect is achieved by affecting the antecedent factors of patients' uncertainty, including patients' fears of physicians' opportunism,

patients' perceived information scarcity, and patients' trust, which are all the traits of principal-agent relationship.

Specifically, the higher the possibility of the physician's opportunism and information scarcity perceived by patients, the greater their uncertainty. Among the antecedents of patient uncertainty, patients' fear of physicians' opportunism has the most significant impact on patients' uncertainty about physicians. By segmenting patients' uncertainty, this research discussed the relationship between patients' uncertainty about the diseases and patients' uncertainty about physicians. The results show that patients' uncertainty about physicians has a significant positive impact on patients' uncertainty about diseases.

In addition, this study also demonstrated the significant role of patients' trust in physicians. Patients' perceived web-based word-of-mouth information about physicians can enhance patients' trust in physicians. Patients' having more information and less fear of physicians' opportunistic behaviors also increases patients' trust. However, from the result of the mediation test, only when the information can increase patients' trust in their physicians, patients' uncertainty about physicians can be reduced; thus, increasing physicians' medical service information can be effective in reducing patients' uncertainty about their physicians. Patients' trust in their physicians fully mediates the relationship between their perception of information scarcity about the physicians' medical service and their uncertainty about their physicians.

Theoretical Contributions

First, on the basis the principal-agent theory and from the perspective of reducing patient uncertainty, this study is the first to explore the influence mechanism of web-based disease-related information quality and web-based word-of-mouth information received by patients on patients' uncertainty. It is to be noted that information can reduce uncertainty, but the mechanism of how information reduces uncertainty is not clear. Therefore, we propose our *uncertainty mitigators–uncertainty antecedents–uncertainty* framework to explore the mechanism. On the basis of the URT, web-based information quality and web-based word-of-mouth information of physicians effectively reduce the antecedents of patients' uncertainty, including perceived information scarcity, fears of physicians' opportunism, and trust. Thus, patients' uncertainty about the disease and the physician are reduced.

Second, this study enriches the literature on patients' uncertainty by classifying patients' uncertainties into patients' uncertainty about the diseases and patients' uncertainty about their physicians. Following the classification of consumers' uncertainty about sellers and products by Dimoka et al [15], we also found that patients' uncertainty about diseases and physicians should be distinguished. In particular, we explored the role of patients' uncertainty about physicians, which has been rarely studied in the existing literature. Reducing patients' uncertainty about their physicians can further reduce their uncertainty about diseases.

Third, this study emphasizes the significant role of patients' trust. As an important factor in principal-agent relationships, trust is the most valuable aspect [43]. We also found that without trust, just increasing patients' information does not help reduce their uncertainty about their physicians. This result further supports the fact that building trust is crucial to address the principal–agent problem.

Practical Contributions

First, this study found that the better the web-based word-of-mouth information of a physician and information quality obtained by patients, the better the reduction in patients'

uncertainty. Therefore, for physicians in the internet era, attention should be paid to the role of web-based health information. More authoritative, more reliable, and higher-quality web-based platforms should be provided to meet patients' demands for health information. In addition, physicians should encourage offline patients to participate in web-based word-of-mouth evaluations, maintain their own web-based word-of-mouth information, and provide more information about their services to potential patients [81]. Web-based word-of-mouth information can reach a wider audience and has a greater impact than offline word-of-mouth information. Web-based word-of-mouth information can effectively nudge physicians to improve their service quality and help patients acquire relevant information about physicians, thereby reducing patients' uncertainty [39].

Second, from the full mediator role of trust, web-based information is effective only when this information can help build patients' trust in their physicians. This suggests that web-based platforms that provide information (ie, web-based word-of-mouth information about physicians) should strictly check the quality of the information. More importantly, platforms can provide some cues to inform patients that the information is trustworthy, such as third-party certifications and guarantees. Only when patients can trust their physicians through this information can it help reduce their uncertainties.

Limitations and Future Directions

First, as for the sample composition, there are 3 prerequisites for this study. Only those who might have a certain disease and had seen the physicians offline within the past 3 months, read the web-based word-of-mouth information about physicians, and engaged in web-based disease information search behaviors were eligible, which resulted in a large overrepresentation of younger people in our sample. More than 90.5% (305/337) of our respondents were aged <40 years, so the sample had possible self-selection bias and a bias of young age. Second, regarding the collection time of data, the research data were collected in May 2020 after the COVID-19 epidemic in China. The external validity of the results may be jeopardized. Then, this study only considered the influence mechanism of web-based word-of-mouth information about physicians on offline patients' trust. Future studies can further consider the situation of web-based health consultation and investigate the possible differences in web-based health information on the physician-patient relationship in different channels. Moreover, because the focus of this study is the information about diseases and physicians, the respondents' health status is controlled, and the result shows that respondents' perception of their health status influences their uncertainty. Future studies can further discuss and explain the effect of health status. Finally, the study data were cross-sectional subjective data, which were provided by the same subjects at the same time, and future studies can use longitudinal analysis or experiments to better test the causal relationships in the model.

Acknowledgments

This work was supported by a grant from the National Natural Science Foundation of China (grants 71771219 and 72071213).

Conflicts of Interest

None declared.

Multimedia Appendix 1

The procedure and results of common method variance.

[[DOCX File , 425 KB - jmir_v24i3e25275_app1.docx](#)]

References

1. Siering M, Jan M. Credence goods and online product reviews: an exploration of the product type concept in the social commerce era. In: Proceedings of the Nineteenth Americas Conference on Information Systems. 2013 Presented at: Proceedings of the Nineteenth Americas Conference on Information Systems; Aug 15-17, 2013; Chicago, Illinois.
2. Ngamvichaikit A, Beise-Zee R. Customer preference for decision authority in credence services. *Managing Service Q* 2014 May 6;24(3):274-229. [doi: [10.1108/msq-03-2013-0033](#)]
3. Mishel MH. Uncertainty in illness. *Image J Nurs Sch* 1988;20(4):225-232. [doi: [10.1111/j.1547-5069.1988.tb00082.x](#)] [Medline: [3203947](#)]
4. Mishel MH. Perceived uncertainty and stress in illness. *Res Nurs Health* 1984 Sep;7(3):163-171. [doi: [10.1002/nur.4770070304](#)] [Medline: [6567948](#)]
5. Neville KL. Uncertainty in illness. An integrative review. *Orthop Nurs* 2003;22(3):206-214. [doi: [10.1097/00006416-200305000-00009](#)] [Medline: [12803150](#)]
6. Mishel M, Braden C. Finding meaning: antecedents of uncertainty in illness. *Nursing Res* 1988 Mar;37(2):98-103. [doi: [10.1097/00006199-198803000-00009](#)]
7. Kirmani A, Rao A. No pain, no gain: a critical review of the literature on signaling unobservable product quality. *J Mark* 2018 Oct 10;64(2):66-79. [doi: [10.1509/jmkg.64.2.66.18000](#)]
8. Benjamins J, Haveman-Nies A, Gunnink M, Goudkuil A, de Vet E. How the use of a patient-accessible health record contributes to patient-centered care: scoping review. *J Med Internet Res* 2021 Jan 11;23(1):e17655 [FREE Full text] [doi: [10.2196/17655](#)] [Medline: [33427683](#)]
9. Mesko B, Györfy Z. The rise of the empowered physician in the digital health era: viewpoint. *J Med Internet Res* 2019 Mar 26;21(3):e12490 [FREE Full text] [doi: [10.2196/12490](#)] [Medline: [30912758](#)]
10. Jacobson P. Empowering the physician-patient relationship: the effect of the Internet. *Partnership* 2007 May 29;2(1):1-13. [doi: [10.21083/partnership.v2i1.244](#)]
11. Wong D, Cheung MK. Online health information seeking and eHealth literacy among patients attending a primary care clinic in Hong Kong: a cross-sectional survey. *J Med Internet Res* 2019 Mar 27;21(3):e10831 [FREE Full text] [doi: [10.2196/10831](#)] [Medline: [30916666](#)]
12. Hedges L, Couey C. How patients use online reviews. *Software Advice*. URL: <https://resourceshub.wpengine.com/resources/how-patients-use-online-reviews/> [accessed 2022-02-07]
13. Sbaifi L, Walton J, Blenkinsopp J, Walton G. Information overload in emergency medicine physicians: a multisite case study exploring the causes, impact, and solutions in four north England national health service trusts. *J Med Internet Res* 2020 Jul 27;22(7):e19126 [FREE Full text] [doi: [10.2196/19126](#)] [Medline: [32716313](#)]
14. Zhang H, Zhang R, Lu X, Zhu X. Impact of personal trust tendency on patient compliance based on internet health information seeking. *Telemed J E Health* 2020 Mar;26(3):294-303. [doi: [10.1089/tmj.2018.0296](#)] [Medline: [31045486](#)]
15. Dimoka A, Hong Y, Pavlou PA. On product uncertainty in online markets: theory and evidence. *MIS Quarterly* 2012 Jun;36(2):395-426. [doi: [10.2307/41703461](#)]
16. Eisenhardt K. Agency theory: an assessment and review. *Acad Manag Rev* 1989 Jan 1;14(1):57-74. [doi: [10.5465/amr.1989.4279003](#)]
17. Akerlof GA. The market for "lemons": quality uncertainty and the market mechanism. *Uncertainty Econ* 1978;235,237-235,251. [doi: [10.1016/B978-0-12-214850-7.50022-X](#)]
18. Rothschild M, Stiglitz J. Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *Quarterly J Econ* 1976 Nov;90(4):629-649. [doi: [10.2307/1885326](#)]
19. Scott A, Vick S. Patients, doctors and contracts: an application of principal-agent theory to the doctor-patient relationship. *Scottish J Political Economy* 1999 May;46(2):111-134. [doi: [10.1111/1467-9485.00124](#)]
20. Langer A, Schröder-Bäck P, Brink A, Eurich J. The agency problem and medical acting: an example of applying economic theory to medical ethics. *Med Health Care Philos* 2009 Mar;12(1):99-108. [doi: [10.1007/s11019-008-9138-y](#)] [Medline: [18470634](#)]
21. Blomqvist A. The doctor as double agent: information asymmetry, health insurance, and medical care. *J Health Econ* 1991 Jan;10(4):411-432. [doi: [10.1016/0167-6296\(91\)90023-g](#)]
22. Suri R, Kohli C, Monroe KB. The effects of perceived scarcity on consumers' processing of price information. *J Acad Mark Sci* 2007 Feb 23;35(1):89-100. [doi: [10.1007/s11747-006-0008-y](#)]

23. Wells JD, Valacich JS, Hess TJ. What signal are you sending? How website quality influences perceptions of product quality and purchase intentions. *MIS Quarterly* 2011;35(2):373-396. [doi: [10.2307/23044048](https://doi.org/10.2307/23044048)]
24. Pavlou PA, Liang H, Xue Y. Understanding and mitigating uncertainty in online exchange relationships: a principal-agent perspective. *MIS Quarterly* 2007;31(1):105-136. [doi: [10.2307/25148783](https://doi.org/10.2307/25148783)]
25. Kohli R, Kettinger WJ. Informing the clan: controlling physicians' costs and outcomes. *MIS Quarterly* 2004;28(3):363-394. [doi: [10.2307/25148644](https://doi.org/10.2307/25148644)]
26. Skinner H, Biscope S, Poland B, Goldberg E. How adolescents use technology for health information: implications for health professionals from focus group studies. *J Med Internet Res* 2003 Dec 18;5(4):e32 [FREE Full text] [doi: [10.2196/jmir.5.4.e32](https://doi.org/10.2196/jmir.5.4.e32)] [Medline: [14713660](https://pubmed.ncbi.nlm.nih.gov/14713660/)]
27. Zahedi FM, Song J. Dynamics of trust revision: using health infomediaries. *J Manag Inf Syst* 2014 Dec 08;24(4):225-248. [doi: [10.2753/mis0742-1222240409](https://doi.org/10.2753/mis0742-1222240409)]
28. Qahri-Saremi H, Montazemi AR. Factors affecting the adoption of an electronic word of mouth message: a meta-analysis. *J Manag Inf Syst* 2019 Aug 04;36(3):969-1001. [doi: [10.1080/07421222.2019.1628936](https://doi.org/10.1080/07421222.2019.1628936)]
29. Manes E, Tchetchik A. The role of electronic word of mouth in reducing information asymmetry: an empirical investigation of online hotel booking. *J Business Res* 2018 Apr;85:185-196. [doi: [10.1016/j.jbusres.2017.12.019](https://doi.org/10.1016/j.jbusres.2017.12.019)]
30. Wang X, Guo J, Wu Y, Liu N. Emotion as signal of product quality. *Internet Res* 2019 Oct 22;30(2):463-485. [doi: [10.1108/intr-09-2018-0415](https://doi.org/10.1108/intr-09-2018-0415)]
31. Zhou W, Duan W. An empirical study of how third-party websites influence the feedback mechanism between online Word-of-Mouth and retail sales. *Decision Support Syst* 2015 Aug;76:14-23. [doi: [10.1016/j.dss.2015.03.010](https://doi.org/10.1016/j.dss.2015.03.010)]
32. Gao G, Greenwood BN, Agarwal R, McCullough JS. Vocal minority and silent majority: how do online ratings reflect population perceptions of quality? *MIS Q* 2015 Mar 3;39(3):565-589. [doi: [10.25300/misq/2015/39.3.03](https://doi.org/10.25300/misq/2015/39.3.03)]
33. Jones K, Leonard LN. Factors influencing buyer's trust in consumer-to-consumer e commerce. *J Comput Inf Syst* 2015 Dec 10;54(4):71-79. [doi: [10.1080/08874417.2014.11645724](https://doi.org/10.1080/08874417.2014.11645724)]
34. Hu Y, Zhang Z. Patient education - A route to improved patient experience in Chinese hospitals? *Patient Educ Couns* 2015 Jun 23. [doi: [10.1016/j.pec.2015.06.015](https://doi.org/10.1016/j.pec.2015.06.015)] [Medline: [26162956](https://pubmed.ncbi.nlm.nih.gov/26162956/)]
35. Huang S, Xia C, Wang D, Tan J. Unpopular perspective of doctor-patient relationship: study on doctors' opportunistic behavior. *Chinese Med Ethics* 2017;30(9):1094,1097-1094,1121.
36. He AJ. The doctor-patient relationship, defensive medicine and overprescription in Chinese public hospitals: evidence from a cross-sectional survey in Shenzhen city. *Soc Sci Med* 2014 Dec;123:64-71. [doi: [10.1016/j.socscimed.2014.10.055](https://doi.org/10.1016/j.socscimed.2014.10.055)] [Medline: [25462606](https://pubmed.ncbi.nlm.nih.gov/25462606/)]
37. Tang L, Guan M. Rise of health consumerism in China and its effects on physicians' professional identity and the physician-patient relationship and communication. *Health Commun* 2018 May;33(5):636-642. [doi: [10.1080/10410236.2017.1290015](https://doi.org/10.1080/10410236.2017.1290015)] [Medline: [28281791](https://pubmed.ncbi.nlm.nih.gov/28281791/)]
38. Leisen B, Hyman MR. Antecedents and consequences of trust in a service provider. *J Business Res* 2004 Sep;57(9):990-999. [doi: [10.1016/s0148-2963\(02\)00343-0](https://doi.org/10.1016/s0148-2963(02)00343-0)]
39. Bates DW, Gawande AA. The impact of the internet on quality measurement. *Health Aff (Millwood)* 2000;19(6):104-114. [doi: [10.1377/hlthaff.19.6.104](https://doi.org/10.1377/hlthaff.19.6.104)] [Medline: [11192393](https://pubmed.ncbi.nlm.nih.gov/11192393/)]
40. Pauwels K, Aksehirli Z, Lackman A. Like the ad or the brand? Marketing stimulates different electronic word-of-mouth content to drive online and offline performance. *Int J Res Market* 2016 Sep;33(3):639-655. [doi: [10.1016/j.ijresmar.2016.01.005](https://doi.org/10.1016/j.ijresmar.2016.01.005)]
41. Lee V. Transparency and trust — online patient reviews of physicians. *N Engl J Med* 2017 Jan 19;376(3):197-199. [doi: [10.1056/nejmp1610136](https://doi.org/10.1056/nejmp1610136)]
42. Gao G, Greenwood BN, Agarwal R, McCullough JS. Vocal minority and silent majority: how do online ratings reflect population perceptions of quality? *MIS Q* 2015 Mar 3;39(3):565-589. [doi: [10.25300/misq/2015/39.3.03](https://doi.org/10.25300/misq/2015/39.3.03)]
43. Mayer RC, Davis JH, Schoorman FD. An integrative model of organizational trust. *Acad Manag Rev* 1995 Jul;20(3):709-734. [doi: [10.5465/amr.1995.9508080335](https://doi.org/10.5465/amr.1995.9508080335)]
44. Lewis JD, Weigert A. Trust as a social reality. *Social Forces* 1985 Jun;63(4):967-985. [doi: [10.2307/2578601](https://doi.org/10.2307/2578601)]
45. Liang H, Laoethakul K, Lloyd SJ, Xue Y. Information systems and health care-I: trust, uncertainty, and online prescription filling. *Commun Assoc Inf Syst* 2005;15:41-60. [doi: [10.17705/1cais.01502](https://doi.org/10.17705/1cais.01502)]
46. Cummings L. The organizational trust inventory (OTI): development and validation. In: *Trust in Organizations: Frontiers of Theory and Research*. Thousand Oaks, California: SAGE Publications; 1996.
47. Mohd Kassim N, Kader Mohammed Ahmed Abdulla A. The influence of attraction on internet banking: an extension to the trust - relationship commitment model. *Intl J Bank Marketing* 2006 Oct;24(6):424-442. [doi: [10.1108/02652320610701744](https://doi.org/10.1108/02652320610701744)]
48. Zhou W, Duan W. Do professional reviews affect online user choices through user reviews? An empirical study. *J Manag Inf Syst* 2016 Jun 17;33(1):202-228. [doi: [10.1080/07421222.2016.1172460](https://doi.org/10.1080/07421222.2016.1172460)]
49. Ye Q, Law R, Gu B, Chen W. The influence of user-generated content on traveler behavior: an empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Comput Human Behav* 2011 Mar;27(2):634-639. [doi: [10.1016/j.chb.2010.04.014](https://doi.org/10.1016/j.chb.2010.04.014)]

50. Han X, Qu J, Zhang T. Exploring the impact of review valence, disease risk, and trust on patient choice based on online physician reviews. *Telematics Informatics* 2019 Dec;45:101276. [doi: [10.1016/j.tele.2019.101276](https://doi.org/10.1016/j.tele.2019.101276)]
51. Lu W, Wu H. How online reviews and services affect physician outpatient visits: content analysis of evidence from two online health care communities. *JMIR Med Inform* 2019 Dec 02;7(4):e16185 [FREE Full text] [doi: [10.2196/16185](https://doi.org/10.2196/16185)] [Medline: [31789597](https://pubmed.ncbi.nlm.nih.gov/31789597/)]
52. Srivastava SC, Chandra S. Social presence in virtual world collaboration: an uncertainty reduction perspective using a mixed methods approach. *MIS Q* 2018 Mar 3;42(3):779-803. [doi: [10.25300/misq/2018/11914](https://doi.org/10.25300/misq/2018/11914)]
53. Venkatesh V, Thong JY, Chan FK, Hu PJ. Managing citizens' uncertainty in e-government services: the mediating and moderating roles of transparency and trust. *Inf Syst Res* 2016 Mar;27(1):87-111. [doi: [10.1287/isre.2015.0612](https://doi.org/10.1287/isre.2015.0612)]
54. Ogden J, Fuks K, Gardner M, Johnson S, McLean M, Martin P, et al. Doctors expressions of uncertainty and patient confidence. *Patient Educ Counseling* 2002 Oct;48(2):171-176. [doi: [10.1016/s0738-3991\(02\)00020-4](https://doi.org/10.1016/s0738-3991(02)00020-4)]
55. Zhang Y. Uncertainty in illness: theory review, application, and extension. *Oncol Nurs Forum* 2017 Nov 1;44(6):645-649. [doi: [10.1188/17.onf.645-649](https://doi.org/10.1188/17.onf.645-649)]
56. Yoon P. An agency theory perspective on physician interactions with the pharmaceutical industry. *Acad Manag Annual Meeting Proc* 2004 Aug;2004(1):D1-D6. [doi: [10.5465/ambpp.2004.13863045](https://doi.org/10.5465/ambpp.2004.13863045)]
57. *Multivariate Data Analysis (4th Ed.): With Readings*. United States: Prentice-Hall, Inc; 1995.
58. Lewis BR, Templeton GF, Byrd TA. A methodology for construct development in MIS research. *Eur J Inf Syst* 2017 Dec 19;14(4):388-400. [doi: [10.1057/palgrave.ejis.3000552](https://doi.org/10.1057/palgrave.ejis.3000552)]
59. Zahedi F, Song J. Dynamics of trust revision: using health infomediaries. *J Manag Inf Syst* 2014 Dec 08;24(4):225-248. [doi: [10.2753/mis0742-1222240409](https://doi.org/10.2753/mis0742-1222240409)]
60. Collins CJ, Stevens CK. The relationship between early recruitment-related activities and the application decisions of new labor-market entrants: a brand equity approach to recruitment. *J Appl Psychol* 2002 Dec;87(6):1121-1133. [doi: [10.1037/0021-9010.87.6.1121](https://doi.org/10.1037/0021-9010.87.6.1121)] [Medline: [12558218](https://pubmed.ncbi.nlm.nih.gov/12558218/)]
61. McKnight DH, Choudhury V, Kacmar C. Developing and validating trust measures for e-commerce: an integrative typology. *Inf Syst Res* 2002 Sep;13(3):334-359. [doi: [10.1287/isre.13.3.334.81](https://doi.org/10.1287/isre.13.3.334.81)]
62. Zhou Y, Kankanhalli A, Yang Z, Lei J. Expectations of patient-centred care: investigating IS-related and other antecedents. *Inf Manag* 2017 Jul;54(5):583-598. [doi: [10.1016/j.im.2016.11.009](https://doi.org/10.1016/j.im.2016.11.009)]
63. Torkzadeh G, Dhillon G. Measuring factors that influence the success of internet commerce. *Inf Syst Res* 2002 Jun;13(2):187-204. [doi: [10.1287/isre.13.2.187.87](https://doi.org/10.1287/isre.13.2.187.87)]
64. Kim DJ, Ferrin DL, Rao HR. A trust-based consumer decision-making model in electronic commerce: the role of trust, perceived risk, and their antecedents. *Decision Support Syst* 2008 Jan;44(2):544-564. [doi: [10.1016/j.dss.2007.07.001](https://doi.org/10.1016/j.dss.2007.07.001)]
65. Guo S, Guo X, Fang Y, Vogel D. How doctors gain social and economic returns in online health-care communities: a professional capital perspective. *J Manag Inf Syst* 2017 Aug 17;34(2):487-519. [doi: [10.1080/07421222.2017.1334480](https://doi.org/10.1080/07421222.2017.1334480)]
66. Urbach N, Ahlemann F. Structural equation modeling in information systems research using partial least squares. *J Inf Technol Theory Application* 2010 Jun;11(2):5-40.
67. 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](https://doi.org/10.1037/0021-9010.88.5.879)] [Medline: [14516251](https://pubmed.ncbi.nlm.nih.gov/14516251/)]
68. Podsakoff PM, Organ DW. Self-reports in organizational research: problems and prospects. *J Manag* 2016 Sep 04;12(4):531-544. [doi: [10.1177/014920638601200408](https://doi.org/10.1177/014920638601200408)]
69. Liang H, Saraf N, Hu Q, Xue Y. Assimilation of enterprise systems: the effect of institutional pressures and the mediating role of top management. *MIS Q* 2007;31(1):59-87. [doi: [10.2307/25148781](https://doi.org/10.2307/25148781)]
70. Williams LJ, Edwards JR, Vandenberg RJ. Recent advances in causal modeling methods for organizational and management research. *J Manag* 2003 Dec 01;29(6):903-936. [doi: [10.1016/s0149-2063_03_00084-9](https://doi.org/10.1016/s0149-2063_03_00084-9)]
71. Wetzels M, Odekerken-Schröder G, van Oppen C. Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration. *MIS Q* 2009;33(1):177-195. [doi: [10.2307/20650284](https://doi.org/10.2307/20650284)]
72. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. *J Market Theory Pract* 2014 Dec 08;19(2):139-152. [doi: [10.2753/mtp1069-6679190202](https://doi.org/10.2753/mtp1069-6679190202)]
73. Straub D, Gefen D. Validation guidelines for IS positivist research. *Commun Assoc Inf Syst* 2004;13:380-427. [doi: [10.17705/1cais.01324](https://doi.org/10.17705/1cais.01324)]
74. Werts C, Linn R, Jöreskog K. Intraclass reliability estimates: testing structural assumptions. *Educ Psychol Measur* 2016 Jul 02;34(1):25-33. [doi: [10.1177/001316447403400104](https://doi.org/10.1177/001316447403400104)]
75. Bernstein I, Nunnally J. The assessment of reliability. *Psychometric Theory* 1994:248-292.
76. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. *J Market Res* 2018 Nov 28;18(1):39-50. [doi: [10.1177/002224378101800104](https://doi.org/10.1177/002224378101800104)]
77. Hair JF, Hult GT, Ringle CM, Sarstedt M. *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. New York: Sage publications; 2016.
78. Sobel ME. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodol* 1982;13:290-312. [doi: [10.2307/270723](https://doi.org/10.2307/270723)]

79. Lowry PB, Gaskin J. Partial Least Squares (PLS) Structural Equation Modeling (SEM) for building and testing behavioral causal theory: when to choose it and how to use it. *IEEE Trans Profess Commun* 2014 Jun;57(2):123-146. [doi: [10.1109/tpc.2014.2312452](https://doi.org/10.1109/tpc.2014.2312452)]
80. Sobel test calculator for the significance of mediation software. Free Statistics Calculators. URL: <https://www.danielsoper.com/statcalc> [accessed 2022-02-07]
81. Kordzadeh N. Investigating bias in the online physician reviews published on healthcare organizations' websites. *Decision Support Systems* 2019 Mar;118:70-82. [doi: [10.1016/j.dss.2018.12.007](https://doi.org/10.1016/j.dss.2018.12.007)]

Abbreviations

AVE: average variance extracted
CR: composite reliability
PLS: partial least squares
URT: uncertainty reduction theory

Edited by G Eysenbach; submitted 26.10.20; peer-reviewed by S Bidmon, B Giesser; comments to author 21.12.20; revised version received 08.05.21; accepted 27.12.21; published 11.03.22.

Please cite as:

Dong W, Lei X, Liu Y

The Mediating Role of Patients' Trust Between Web-Based Health Information Seeking and Patients' Uncertainty in China: Cross-sectional Web-Based Survey

J Med Internet Res 2022;24(3):e25275

URL: <https://www.jmir.org/2022/3/e25275>

doi: [10.2196/25275](https://doi.org/10.2196/25275)

PMID: [35275074](https://pubmed.ncbi.nlm.nih.gov/35275074/)

©Wei Dong, Xiangxi Lei, Yongmei Liu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 11.03.2022. 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.

Original Paper

Social Media Images as an Emerging Tool to Monitor Adherence to COVID-19 Public Health Guidelines: Content Analysis

Sean D Young^{1,2}, MSc, PhD; Qingpeng Zhang³, PhD; Daniel Dajun Zeng⁴, PhD; Yongcheng Zhan⁵, PhD; William Cumberland⁶, PhD

¹Department of Informatics, University of California Institute for Prediction Technology, University of California, Irvine, Irvine, CA, United States

²Department of Emergency Medicine, University of California, Irvine, Irvine, CA, United States

³School of Data Science, City University of Hong Kong, Kowloon, Hong Kong, Hong Kong

⁴Institute of Automation, Chinese Academy of Sciences, Beijing, China

⁵Department of Information Systems, California Polytechnic State University, San Luis Obispo, CA, United States

⁶Department of Biostatistics, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:

Sean D Young, MSc, PhD

Department of Informatics

University of California Institute for Prediction Technology

University of California, Irvine

6091 Bren Hall

Irvine, CA, 92617

United States

Phone: 1 650 223 5448

Email: syoung5@hs.uci.edu

Abstract

Background: Innovative surveillance methods are needed to assess adherence to COVID-19 recommendations, especially methods that can provide near real-time or highly geographically targeted data. Use of location-based social media image data (eg, Instagram images) is one possible approach that could be explored to address this problem.

Objective: We seek to evaluate whether publicly available near real-time social media images might be used to monitor COVID-19 health policy adherence.

Methods: We collected a sample of 43,487 Instagram images in New York from February 7 to April 11, 2020, from the following location hashtags: #Centralpark (n=20,937), #Brooklyn Bridge (n=14,875), and #Timesquare (n=7675). After manually reviewing images for accuracy, we counted and recorded the frequency of valid daily posts at each of these hashtag locations over time, as well as rated and counted whether the individuals in the pictures at these location hashtags were social distancing (ie, whether the individuals in the images appeared to be distanced from others vs next to or touching each other). We analyzed the number of images posted over time and the correlation between trends among hashtag locations.

Results: We found a statistically significant decline in the number of posts over time across all regions, with an approximate decline of 17% across each site ($P < .001$). We found a positive correlation between hashtags (#Centralpark and #Brooklynbridge: $r = 0.40$; #BrooklynBridge and #Timesquare: $r = 0.41$; and #Timesquare and #Centralpark: $r = 0.33$; $P < .001$ for all correlations). The logistic regression analysis showed a mild statistically significant increase in the proportion of posts over time with people appearing to be social distancing at Central Park ($P = .004$) and Brooklyn Bridge ($P = .02$) but not for Times Square ($P = .16$).

Conclusions: Results suggest the potential of using location-based social media image data as a method for surveillance of COVID-19 health policy adherence. Future studies should further explore the implementation and ethical issues associated with this approach.

(*J Med Internet Res* 2022;24(3):e24787) doi:[10.2196/24787](https://doi.org/10.2196/24787)

KEYWORDS

internet; social media; health informatics; tool; monitor; adherence; COVID-19; public health; guidelines; content analysis; policy

Introduction

Innovative surveillance methods are needed to assess adherence to COVID-19 recommendations [1], especially methods that can provide near real-time or highly geographically targeted data [2,3]. Social media, phone mobility data, and digital tracing apps have been discussed as potential data sources to use to better understand and track COVID-19-related behaviors and policy adherence [4-8]. However, no known COVID-19 or other research has examined whether social media images posted on a location hashtag (eg, #Centralpark) might inform regional surveillance and intervention efforts. Understanding these trends in local adherence to emergency public health orders could help inform public health and clinical needs. Accordingly, in a pilot study, we evaluated whether publicly available images might be used as a low-cost, near real-time method to monitor adherence to COVID-19 health policies.

Methods

Instagram is the most popular photo-sharing application in the United States. It allows users to take pictures of their current activities and environment, and share them with others in real time. Pictures can have location tags where users can post pictures to a specific topic thread (eg, pictures taken in Central Park, New York). Approximately 37% of US adults use Instagram, with 75% of use among those 18-24 years of age, 57% among those 25-29 years of age, 47% among those 30-49 years of age, 23% among those 50-64 years of age, and 8% among people 65 years and older [9].

We built a Python web crawler to collect Instagram images and corresponding user information and metadata. Specifically, the Beautiful Soup package was used to parse the original HTML files. The keywords, with hashtags such as #brooklynbridge, were used to identify relevant data. The images were only collected from public accounts, which led to a sample of 43,487 Instagram images in New York from February 7 to April 11, 2020, during a timeline throughout which New York COVID-19-related public health recommendations shifted from no recommendations (March 1, 2020, the first confirmed case within New York; March 5, 2020, Mayor de Blasio reports that fears should not keep New Yorkers off the subway), to heightened awareness (March 7, 2020, Governor Cuomo declares a state of emergency), to statewide stay-at-home orders for all nonessential activities, including limiting all outdoor activities with a possibility of coming into close contact with others (March 22, 2020). For example, as part of the stay-at-home order, New Yorkers were instructed that all nonessential gatherings of individuals of any size for any reason should be canceled or postponed. They were also informed that individuals should limit outdoor recreational activities to noncontact and avoid activities where they come in close contact with other people [10].

Images were only collected from location hashtags known for nonessential activities or crowded spaces where it would likely be difficult for people to socially distance (#Centralpark; n=20,937), #Brooklyn Bridge (n=14,875), and #Timesquare (n=7675). These data were collected to attempt to describe the

changing response to stay-at-home orders within these locations. We excluded images that might have been posted by the same person by only retaining up to 1 image per day per username. We also excluded images if we were unable to verify location. The final sample of images were manually reviewed by 23 graduate students to attempt to visually verify that the individuals in the pictures were at the hashtag locations (eg, images posted to #Brooklynbridge were excluded if they appeared to be indoors). Intercooder reliability was assessed by having students each label a subset of the sample of pictures to determine consistency. These students then reviewed the labels to resolve inconsistencies. This process occurred on the final sample.

We counted and recorded the frequency of valid daily posts at each of these hashtag locations over time (trend in frequency of posts at each hashtag location), as well as rated and counted whether the individuals in the pictures at these location hashtags were social distancing (ie, whether the individuals in the images appeared to be distanced from others vs next to or touching each other). As an additional metric for assessing the reliability/consistency of findings, we also measured the correlation between the location hashtags in their frequency of posts over time using data from the entire sample of individuals associated with these locations. The graduate students rating the pictures were provided with guidance on how to conduct the rating and reviewed for agreement, including instructing them to estimate social distancing based on individuals in the photos not touching or being directly next to other individuals (ie, not posing with, standing, or sitting next to each other).

We used R software (R Foundation for Statistical Computing) to conduct regressions analyzing the number of images posted over time and to calculate the correlation between trends among hashtag locations (eg, the correlation between frequency of daily images posted to #Centralpark and #Brooklyn Bridge). For illustration purposes, we calculated the average number of images posted before versus after the first New York case (March 1, 2020). Logistic regression was used to estimate changes in the proportion of posts exhibiting social distancing over time. City University of Hong Kong research ethics committee (#2-25-202001-01) and the University of California, Irvine Institutional Review Board approved this study.

Results

The final sample included 37,447 manually verified images: #Centralpark (n=17,761), #Brooklynbridge (n=13,459), and #Timesquare (n=6227). A total of 100 randomly selected images were reviewed by 4 of the 23 labelers to assess intercooder reliability that individuals in the pictures were at the reported hashtag locations ($\kappa=0.64$). We found a statistically significant decline in the number of posts over time across all regions, with an approximate decline of 17% across each site ($P<.001$; Figure 1 and Table 1). We found a positive correlation between hashtags (#Centralpark and #Brooklynbridge: $r=0.40$; #BrooklynBridge and #Timesquare: $r=0.41$; and #Timesquare and #Centralpark: $r=0.33$; $P<.001$ for all correlations). The logistic regression analysis showed a mild statistically significant increase in the proportion of posts over time with people

appearing to be social distancing at Central Park ($P=.004$) and Brooklyn Bridge ($P=.02$) but not for Times Square ($P=.16$).

Figure 1. Number of images posted to the location hashtags over time, New York, 2020. Trendlines are estimated by a third-order polynomial function.

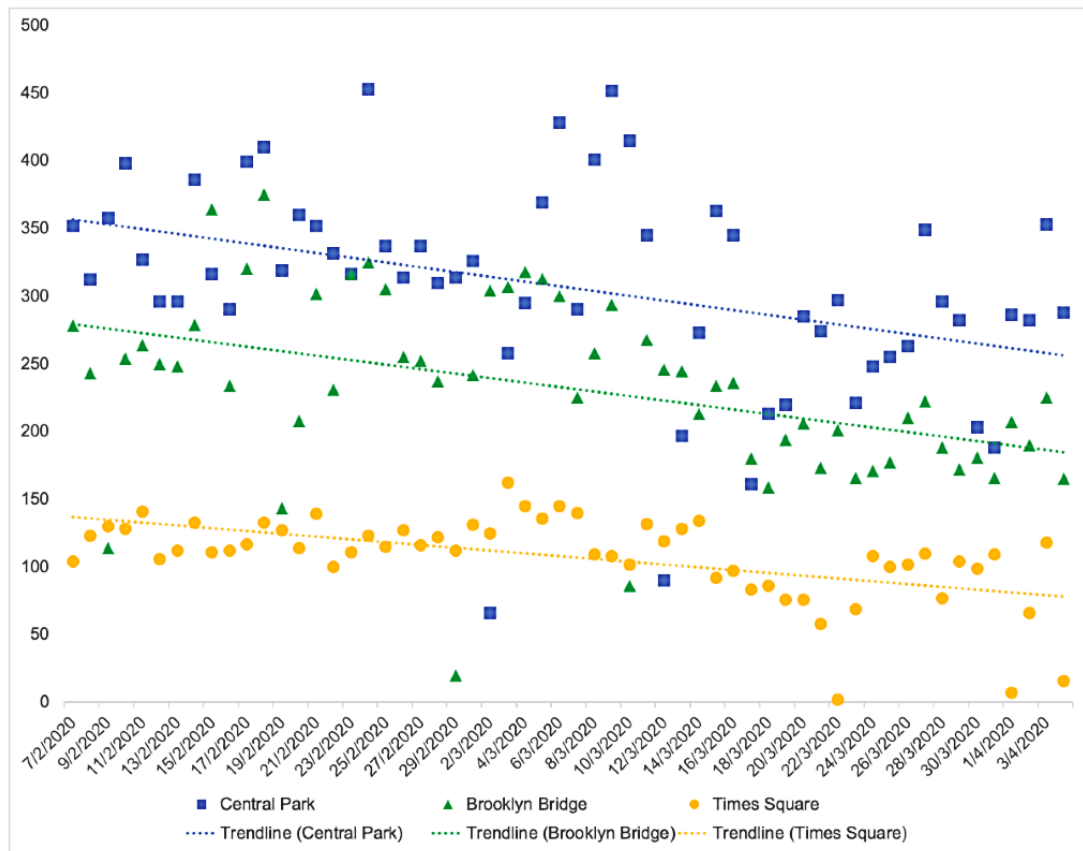


Table 1. Change in mean number of daily Instagram images for each location, New York, 2020.

Location	Number of daily posts before March 1, 2020, mean	Number of daily posts after March 1, 2020, mean (% change)	Slope	95% CI	P value
Central Park	342.8	282.2 (-13.7)	-1.76	-2.89 to -0.62	<.001
Brooklyn Bridge	252.9	218.3 (-17.7)	-1.66	-2.61 to -0.71	<.001
Times Square	119.8	99.2 (-17.2)	-1.03	-1.45 to -0.62	<.001

Discussion

Results suggest that publicly available image data might be incorporated into public health surveillance methodologies as an additional tool for monitoring people’s adherence to public health guidelines, such as for the COVID-19 pandemic. We collected more than 40,000 images throughout a 10-week study period, which provided potential information about people’s locations and adherence to stay-at-home orders. Sample data were a small subset of available images, and for only 1 city, suggesting that this approach could be scaled and automated with artificial intelligence to assist with near real-time regional health surveillance. Although this is a pilot study to explore this new approach for surveillance, it provides an opportunity for future researchers to explore expanding these methods using artificial intelligence and to assess the potential cost-effectiveness of this approach.

There are a number of potential public health and emergency clinical applications of this research. First, health officials might

use social media image data to better understand trends in adherence to COVID-19 prevention and other health policies. Second, using similar methods as other areas of health informatics, social media image data could be analyzed in health prediction models alongside other data sources, including case diagnoses, health services use, and demographic information, to learn how social media images might predict future COVID-19 cases within a specific region or county [11,12]. Finally, by providing data on where, when, how, and who are adhering or not adhering to health recommendations within a specific region, these data might help inform both the need for and ability to craft education and behavior change campaigns that are tailored to specific demographic or regional audiences [13,14], as well as trends in potential future local emergency department visits related to COVID-19 [15].

Although the number of posts decreased by approximately 17% throughout the study period, on average, 600 posts continued per day after issuance of stay-at-home orders, supporting the continued need for behavior change interventions. Although

we found a correlation between locations, the percentage reduction in images posted was greatest for #Timesquare and least for #Centralpark. This may be because Times Square is primarily a tourist location (and tourist activity substantially diminished), while people may have continued to use Central Park to exercise.

This study was limited by a New York focus and inability to verify the exact location, time of photo, or demographic information (eg, race, sex, or home city) about the users and a biased sample of Instagram users. Future research may help to address these questions by incorporating survey data into the current types of methods. Instagram's younger demographic [9] might help to explain the relatively small reduction in images posted after issuance of stay-at-home orders due to the common desire for independence and reduced risk perception among this age group. In addition, there are limitations with the social distancing measure (eg, it would code household members sitting together in the park having a picnic as not adhering to social distancing, when household members sitting together is appropriate, and we were unable to verify the specific number of feet individuals were from each other). We are also unable to identify why people were in these locations. As restrictions lessened over time to allow certain recreational activities, it would be helpful in future research to develop tools (eg, through

artificial intelligence) that could help to identify the specific types of activities participants are doing to better understand if they were adhering to policy recommendations. Despite being publicly accessible data and social media-based public health surveillance research being supported by government agencies (eg, Centers for Disease Control and Prevention) [12,16-18], privacy and ethics-related concerns need to be further explored before implementation [19]. Relatedly, in the future, it is possible that people would intentionally choose to alter their behavior due to demand characteristics or surveillance efforts [20], including deciding to not post to certain location hashtags if they thought this method might be used for surveillance efforts. However, we believe it is unlikely that people would alter their behavior in this way as people continue to publicly share large amounts of personal health information (eg, sexual risk behaviors and drug use) on social media despite ongoing monitoring of social media and digital data [21].

As states and local health departments continue to issue public health orders, a large number of surveillance tools and approaches are needed to control the growing COVID-19 pandemic. Results from this pilot study suggest that image data might be explored and integrated with traditional epidemiology approaches to help address and better inform local health and emergency department efforts.

Acknowledgments

We wish to thank Dr Lifang Li and Mr Jiannan Ye for cleaning the data and doing preliminary analysis. SDY reports grant funding to University of California, Irvine from the National Institute of Allergy and Infectious Diseases (NIAID) and National Institute on Drug Abuse (NIDA). The funders played no role in the study planning, analysis, or outcomes.

Authors' Contributions

QZ and DDZ had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. YZ collected the data. SDY conceived of the initial idea, advised on analysis, and wrote the first draft of the manuscript. WC advised on analysis and edited the manuscript.

Conflicts of Interest

SDY has received equity in MotiSpark, a company that uses images/videos for health behavior change. SDY has been a previous recipient of gift funding from Facebook (funding was made to the University of California with SDY as PI).

References

1. An approach for monitoring and evaluating community mitigation strategies for COVID-19. Centers for Disease Control and Prevention. 2020 Feb 11. URL: <https://tinyurl.com/mrxd99uu> [accessed 2020-11-04]
2. 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](https://doi.org/10.1001/jama.2020.3151)] [Medline: [32125371](https://pubmed.ncbi.nlm.nih.gov/32125371/)]
3. Smith CD, Mennis J. Incorporating geographic information science and technology in response to the COVID-19 pandemic. *Prev Chronic Dis* 2020 Jul 09;17:E58 [FREE Full text] [doi: [10.5888/pcd17.200246](https://doi.org/10.5888/pcd17.200246)] [Medline: [32644920](https://pubmed.ncbi.nlm.nih.gov/32644920/)]
4. Ting DSW, Carin L, Dzau V, Wong TY. Digital technology and COVID-19. *Nat Med* 2020 Apr;26(4):459-461 [FREE Full text] [doi: [10.1038/s41591-020-0824-5](https://doi.org/10.1038/s41591-020-0824-5)] [Medline: [32284618](https://pubmed.ncbi.nlm.nih.gov/32284618/)]
5. Young SD, Schneider J. Clinical care, research, and telehealth services in the era of social distancing to mitigate COVID-19. *AIDS Behav* 2020 Jul;24(7):2000-2002 [FREE Full text] [doi: [10.1007/s10461-020-02924-z](https://doi.org/10.1007/s10461-020-02924-z)] [Medline: [32440971](https://pubmed.ncbi.nlm.nih.gov/32440971/)]
6. Protocol for assessment of potential risk factors for coronavirus disease 2019 (COVID-19) among health workers in a health care setting. World Health Organization. 2020. URL: https://apps.who.int/iris/bitstream/handle/10665/332071/WHO-2019-nCoV-HCW_risk_factors_protocol-2020.3-eng.pdf [accessed 2020-05-21]
7. Oliver N, Lepri B, Sterly H, Lambiotte R, Deletaille S, De Nadai M, et al. Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle. *Sci Adv* 2020 Jun;6(23):eabc0764 [FREE Full text] [doi: [10.1126/sciadv.abc0764](https://doi.org/10.1126/sciadv.abc0764)] [Medline: [32548274](https://pubmed.ncbi.nlm.nih.gov/32548274/)]

8. Garrett R, Young SD. Online misinformation and vaccine hesitancy. *Transl Behav Med* 2021 Dec 14;11(12):2194-2199 [FREE Full text] [doi: [10.1093/tbm/ibab128](https://doi.org/10.1093/tbm/ibab128)] [Medline: [34529080](https://pubmed.ncbi.nlm.nih.gov/34529080/)]
9. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. URL: <https://tinyurl.com/3un9k66e> [accessed 2020-05-21]
10. New York Forward. URL: <https://coronavirus.health.ny.gov/new-york-state-pause> [accessed 2020-11-04]
11. Young SD, Torrone EA, Urata J, Aral SO. Using search engine data as a tool to predict syphilis. *Epidemiology* 2018 Jul;29(4):574-578 [FREE Full text] [doi: [10.1097/EDE.0000000000000836](https://doi.org/10.1097/EDE.0000000000000836)] [Medline: [29864105](https://pubmed.ncbi.nlm.nih.gov/29864105/)]
12. Aiello AE, Renson A, Zivich PN. Social media- and internet-based disease surveillance for public health. *Annu Rev Public Health* 2020 Apr 02;41:101-118 [FREE Full text] [doi: [10.1146/annurev-publhealth-040119-094402](https://doi.org/10.1146/annurev-publhealth-040119-094402)] [Medline: [31905322](https://pubmed.ncbi.nlm.nih.gov/31905322/)]
13. Young SD, Goldstein NJ. Applying social norms interventions to increase adherence to COVID-19 prevention and control guidelines. *Prev Med* 2021 Apr;145:106424 [FREE Full text] [doi: [10.1016/j.ypmed.2021.106424](https://doi.org/10.1016/j.ypmed.2021.106424)] [Medline: [33440191](https://pubmed.ncbi.nlm.nih.gov/33440191/)]
14. Young S. *Stick with It: A Scientifically Proven Process for Changing Your Life--for Good*. Toronto: HarperCollins; Jun 26, 2018.
15. Christie A, Henley SJ, Mattocks L, Fernando R, Lansky A, Ahmad FB, et al. Decreases in COVID-19 Cases, Emergency Department Visits, Hospital Admissions, and Deaths Among Older Adults Following the Introduction of COVID-19 Vaccine - United States, September 6, 2020-May 1, 2021. *MMWR Morb Mortal Wkly Rep* 2021 Jun 11;70(23):858-864 [FREE Full text] [doi: [10.15585/mmwr.mm7023e2](https://doi.org/10.15585/mmwr.mm7023e2)] [Medline: [34111059](https://pubmed.ncbi.nlm.nih.gov/34111059/)]
16. Young SD, Wang W, Chakravarthy B. Crowdsourced traffic data as an emerging tool to monitor car crashes. *JAMA Surg* 2019 Aug 01;154(8):777-778 [FREE Full text] [doi: [10.1001/jamasurg.2019.1167](https://doi.org/10.1001/jamasurg.2019.1167)] [Medline: [31116356](https://pubmed.ncbi.nlm.nih.gov/31116356/)]
17. Zhang Q, Chai Y, Li X, Young SD, Zhou J. Using internet search data to predict new HIV diagnoses in China: a modelling study. *BMJ Open* 2018 Oct 17;8(10):e018335 [FREE Full text] [doi: [10.1136/bmjopen-2017-018335](https://doi.org/10.1136/bmjopen-2017-018335)] [Medline: [30337302](https://pubmed.ncbi.nlm.nih.gov/30337302/)]
18. Young SD, Mercer N, Weiss RE, Torrone EA, Aral SO. Using social media as a tool to predict syphilis. *Prev Med* 2018 Apr;109:58-61 [FREE Full text] [doi: [10.1016/j.ypmed.2017.12.016](https://doi.org/10.1016/j.ypmed.2017.12.016)] [Medline: [29278678](https://pubmed.ncbi.nlm.nih.gov/29278678/)]
19. Gasser U, Ienca M, Scheibner J, Sleigh J, Vayena E. Digital tools against COVID-19: taxonomy, ethical challenges, and navigation aid. *Lancet Digit Health* 2020 Aug;2(8):e425-e434 [FREE Full text] [doi: [10.1016/S2589-7500\(20\)30137-0](https://doi.org/10.1016/S2589-7500(20)30137-0)] [Medline: [32835200](https://pubmed.ncbi.nlm.nih.gov/32835200/)]
20. Young SD, Adelstein BD, Ellis SR. Demand characteristics in assessing motion sickness in a virtual environment: or does taking a motion sickness questionnaire make you sick? *IEEE Trans Vis Comput Graph* 2007;13(3):422-428. [doi: [10.1109/TVCG.2007.1041](https://doi.org/10.1109/TVCG.2007.1041)] [Medline: [17356210](https://pubmed.ncbi.nlm.nih.gov/17356210/)]
21. Garrett R, Smith J, Young SD. A review of social media technologies across the global HIV care continuum. *Curr Opin Psychol* 2016 Jun 01;9:56-66 [FREE Full text] [doi: [10.1016/j.copsyc.2015.10.024](https://doi.org/10.1016/j.copsyc.2015.10.024)] [Medline: [26925455](https://pubmed.ncbi.nlm.nih.gov/26925455/)]

Edited by C Basch; submitted 05.10.20; peer-reviewed by D Bychkov, K McCausland; comments to author 23.10.20; revised version received 24.11.20; accepted 21.12.21; published 03.03.22.

Please cite as:

Young SD, Zhang Q, Zeng DD, Zhan Y, Cumberland W

Social Media Images as an Emerging Tool to Monitor Adherence to COVID-19 Public Health Guidelines: Content Analysis

J Med Internet Res 2022;24(3):e24787

URL: <https://www.jmir.org/2022/3/e24787>

doi: [10.2196/24787](https://doi.org/10.2196/24787)

PMID: [34995205](https://pubmed.ncbi.nlm.nih.gov/34995205/)

©Sean D Young, Qingpeng Zhang, Daniel Dajun Zeng, Yongcheng Zhan, William Cumberland. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org/>), 03.03.2022. 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.

Original Paper

An mHealth Intervention to Improve Medication Adherence and Health Outcomes Among Patients With Coronary Heart Disease: Randomized Controlled Trial

Zhao Ni¹, BMedSci, RN, PhD; Bei Wu², PhD; Qing Yang³, PhD; Lijing L Yan⁴, MPH, PhD; Changqing Liu⁵, RN, PhD; Ryan J Shaw³, RN, PhD

¹School of Medicine, Yale University, New Haven, CT, United States

²Rory Meyers College of Nursing, New York University, New York, NY, United States

³School of Nursing, Duke University, Durham, NC, United States

⁴Global Health Research Center, Duke Kunshan University, Kunshan, China

⁵West China Hospital, Sichuan University, Chengdu, China

Corresponding Author:

Zhao Ni, BMedSci, RN, PhD

School of Medicine

Yale University

135 College Street

New Haven, CT, 06510

United States

Phone: 1 646 617 2232

Email: zhao.ni@yale.edu

Abstract

Background: The treatment of many chronic illnesses involves long-term pharmaceutical therapy, but it is an ongoing challenge to find effective ways to improve medication adherence to promote good health outcomes. Cardioprotective medications can prevent the enlargement of harmful clots, cardiovascular symptoms, and poor therapeutic outcomes, such as uncontrolled high blood pressure and hyperlipidemia, for patients with coronary heart disease. Poor adherence to cardioprotective medications, however, has been reported as a global health concern among patients with coronary heart disease, and it is particularly a concern in China.

Objective: This study aimed to evaluate the efficacy of a mobile health (mHealth) intervention using 2 mobile apps to improve medication adherence and health outcomes.

Methods: A randomized, placebo-controlled, 2-arm parallel study was conducted in a major university-affiliated medical center located in Chengdu, China. Participants were recruited by flyers and health care provider referrals. Each participant was observed for 90 days, including a 60-day period of mHealth intervention and a 30-day period of nonintervention follow-up. The study coordinator used WeChat and Message Express to send educational materials and reminders to take medication, respectively. Participants used WeChat to receive both the educational materials and reminders. Participants in the control group only received educational materials. This study received ethics approval from the Duke Health Institutional Review Board (Pro00073395) on May 5, 2018, and was approved by West China Hospital (20170331180037). Recruitment began on May 20, 2018. The pilot phase of this study was registered on June 8, 2016, and the current, larger-scale study was retrospectively registered on January 11, 2021 (ClinicalTrials.gov).

Results: We recruited 230 patients with coronary heart disease. Of these patients, 196 completed the baseline survey and received the intervention. The majority of participants were married (181/196, 92.4%), male (157/196, 80.1%), and lived in urban China (161/196, 82.1%). Participants' average age was 61 years, and half were retired (103/191, 53.9%). More than half the participants (121/196, 61.7%) were prescribed at least 5 medications. The mean decrease in medication nonadherence score was statistically significant at both 60 days ($t_{179}=2.04$, $P=.04$) and 90 days ($t_{155}=3.48$, $P<.001$). Systolic blood pressure and diastolic blood pressure decreased in the experimental group but increased in the control group. The mean decrease in diastolic blood pressure was statistically significant at both 60 days ($t_{160}=2.07$, $P=.04$) and 90 days ($t_{164}=2.21$, $P=.03$). The mean decrease in systolic blood pressure was significantly different in the groups at 90 days ($t_{165}=3.12$, $P=.002$).

Conclusions: The proposed mHealth intervention can improve medication adherence and health outcomes, including systolic blood pressure and diastolic blood pressure.

Trial Registration: ClinicalTrials.gov NCT02793830; <https://clinicaltrials.gov/ct2/show/NCT02793830> and ClinicalTrials.gov NCT04703439; <https://clinicaltrials.gov/ct2/show/NCT04703439>

(*J Med Internet Res* 2022;24(3):e27202) doi:[10.2196/27202](https://doi.org/10.2196/27202)

KEYWORDS

mHealth; medication adherence; coronary disease; blood pressure; China; randomized controlled trial

Introduction

Coronary heart disease (CHD) is a heart and blood vessel disease related to atherosclerosis in which plaque builds up in the walls of the coronary arteries, narrowing the arteries and restricting the flow of blood [1]. CHD is the world's leading cause of death [2], accounting for over one-third of all deaths in individuals over the age of 35 [3]. Approximately 4 out of 5 global deaths caused by CHD occur in low- and middle-income countries [4], and the burden of CHD is anticipated to increase in these countries [5]. By the year 2030, CHD is projected to cause more than 9 million deaths yearly [6]. China is the world's largest low- or middle-income country [7,8], and in China, CHD is the second leading cause of death, accounting for over 1.5 million deaths per year [9,10]. CHD has been an enormous economic burden on Chinese society in terms of health care costs and loss of productivity, and the mortality rate of CHD is increasing [11,12]. The administration of cardioprotective medications is a key treatment modality for CHD and a preventive measure against cardiovascular events [13,14]. Cardioprotective medications, including antiplatelet drugs, beta-blockers, calcium channel blockers, statins, and angiotensin-converting enzyme inhibitors can reduce the mortality rate of CHD [15-17].

In China, however, poor adherence to cardioprotective medications has been reported as a public health concern [18], contributing to increased health care costs and a high mortality rate for cardiovascular diseases [14,19-21]. For example, studies in China have shown that 18.2% to 38% of patients with CHD did not adhere to their statin medications [22,23], and only 49% consistently used beta-blockers after hospital discharge [16,21]. The vast majority of highly qualified health care providers in China practice in large urban hospitals; for serious illness, such as CHD, patients prefer to utilize large hospitals rather than local primary health care clinics [24,25]. Under this treatment-focused health care utilization model, many patients with CHD in China receive prescriptions and medication-related education without a primary care clinician to monitor their medication-taking behavior [20]. Consequently, after being discharged from the hospital, patients often do not receive proper follow-up care or information regarding medication-taking behavior [19,23]; therefore, it is important to develop innovative interventions to address this issue. Mobile health (mHealth) is defined as the use of portable electronic devices with software applications to provide health care services and manage patient information [26-28]. China has 1.3 billion mobile phone users [29]; mobile technology is booming, with 97% of netizens using mobile phones to access the internet [30]. Pilot studies conducted in China have shown that it is feasible and acceptable

to implement an mHealth intervention to improve medication adherence among patients with CHD [19].

After these pilot studies, mobile phone-based mHealth interventions have been conducted in the field of cardiovascular medicine in China [31-34], but none of them has been specifically aimed at improving medication adherence among patients with CHD, nor has any of them tested mHealth interventions that integrate 2 mobile apps. For example, in Shanghai, Dorje et al [31] conducted an mHealth intervention using social media to send participants educational materials to promote cardiac rehabilitation and secondary prevention, but the educational materials were not specifically about medication adherence. Also, reminders to take medication were not a primary intervention in this study; only participants whose blood pressure was outside the target level received reminders. In Guangzhou, Li et al [34] evaluated an mHealth intervention that aimed to improve blood pressure and self-management behavior in people with hypertension, but this intervention was neither specifically designed to improve medication adherence nor focused on patients with CHD. These studies indicate that although using mHealth interventions to improve medication adherence is a promising approach, more evidence is needed to examine the effect of mHealth intervention on improving medication adherence and clinical outcomes among patients with CHD [35]. The aim of this paper was to assess if our mobile phone-based mHealth intervention could improve medication adherence and relevant health outcomes (ie, systolic blood pressure, diastolic blood pressure, and heart rate) among patients with CHD in comparison to a control group that received general educational materials over a period of 60 days. The pilot phase of this study was registered in ClinicalTrials.gov (NCT02793830) on June 8, 2016, and the current, larger-scale study was retrospectively registered (NCT04703439) on January 11, 2021. This was because we wanted to make the study more accessible to the public and also because we believed that this would better reflect the continuousness of our mHealth intervention, from the initial evaluation of its feasibility and acceptability (ie, the pilot phase) to the later evaluation of its efficacy (ie, the larger-scale phase).

Methods

mHealth Technologies

This study used 2 mobile apps: WeChat (Tencent Inc) and Message Express (Bluemobile.zt). WeChat is the most widely used messaging app in China [36]. Message Express (also known as “Xiao Xi Su Di” in Chinese) is a message saving and delivery app that can send personalized messages to a WeChat

account. It can be integrated with WeChat to deliver reminders to take medication from a message library. After saving a library of reminders to Message Express, a clinician can easily choose personalized reminders in Message Express and send them to a participant's WeChat. BB Reminder is a third-party app with similar functions to Message Express that we used in the pilot study, but it became unavailable after April 2018. Thus, in this study, our study coordinator used WeChat and Message Express to send educational materials and reminders to take medication, respectively. Participants used WeChat to receive both educational materials and reminders. At baseline, we assessed the mobile phone literacy of participants by recording the length of time they had been using WeChat, their frequency of WeChat use, and their educational level. To protect participants' data and privacy, we followed the patient privacy policies of West China Hospital and Duke University and used an encrypted cellphone to send messages. Participants' data were stored in a secure database (REDCap, Vanderbilt University). No data were stored in WeChat or Message Express.

Study Design and Participants

This unblinded, 2-arm, parallel randomized controlled trial was conducted between May and December 2018 at the Cardiology Department of West China Hospital, located in Chengdu. West China Hospital is a major university-affiliated hospital that serves more than 10,000 outpatients a day [37]. A total of 230 participants were recruited for this study using flyers and health care provider referrals [19]. The sample size was determined by a statistical power analysis (Gpower 3.1), with an α of 0.05, power of 0.80, and effect size of 0.35. The effect size was calculated using Cohen's criteria based on data from phase II of the pilot study [38,39]. Written consent was obtained from participants, and they were randomly assigned to the experimental group using stratified randomization by gender with a permuted block size of 4. The random allocation sequence was generated using SAS software version 9.4 by the first author (ZN), who also enrolled participants and assigned participants to interventions. Each participant was observed for 90 days, including a 60-day period of mHealth intervention and a 30-day period of nonintervention follow-up. The mHealth intervention was developed and tested in a pilot study with 2 phases, in 2016 and 2017 [19]. In the pilot study, we found it was feasible to conduct an mHealth intervention by integrating 2 apps to remind patients with CHD to take medications. The apps that we chose to use in this study were based on the results of the pilot study. Participants were included in this study if they satisfied the following criteria [19]: (1) they had a medical diagnosis of CHD; (2) they were aged 18 years or older; (3) they had an antihypertensive medication regimen that would last for 90 days

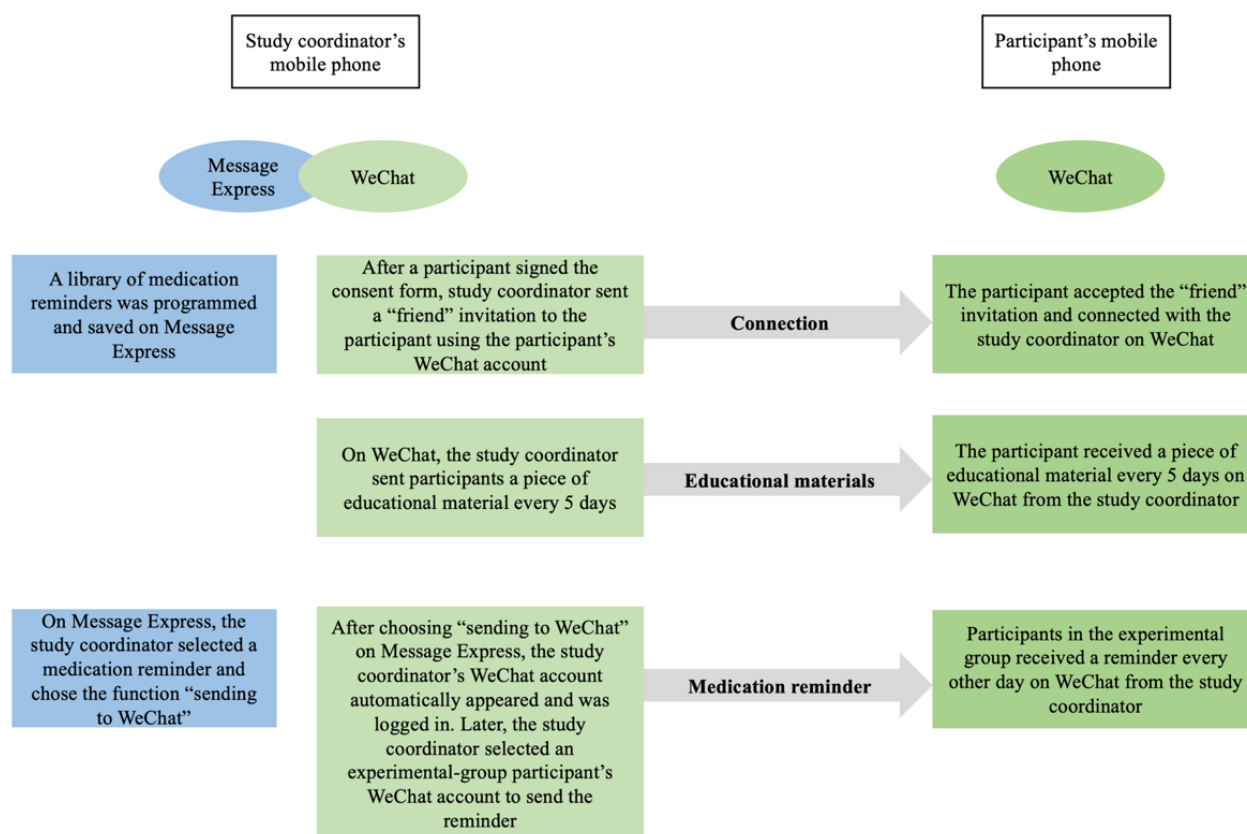
or more after enrollment; (4) they were able to read messages through a mobile phone; (5) they had a mobile phone that could receive messages from WeChat and reminders from Message Express; (6) they were capable of giving consent; and (7) they had an electronic blood pressure cuff to check blood pressure and heart rate.

Ethics Approval

This study received ethical approval from the Duke Health Institutional Review Board (Pro00073395) and West China Hospital (20170331180037).

Interventions

The interventions were refined based on the pilot study [19]. Once a participant completed the baseline questionnaire, they received the allocated intervention. Participants in the experimental group received reminders to take medication and educational materials from Message Express and WeChat, respectively (Figure 1). Participants in the control group only received educational materials from WeChat. The educational materials were sent every 5 days at a random time between 8 AM and 9 AM. The educational materials sent to the 2 groups were different. Materials sent to the experimental group were specifically related to CHD and medication adherence and included information on the function of cardioprotective medications, potential negative consequences of not taking these medications daily, and some common reasons why people with CHD fail to take them. Materials sent to the control group contained general medical information that was not specifically related to CHD or medication adherence. For example, some of the materials were related to population aging in China and age-related health problems, such as hearing loss. All educational materials were retrieved from the website of the World Health Organization [40,41] and were screened by a cardiologist and a nurse to ensure their accuracy before being sent to participants. In addition to educational materials, participants in the experimental group received reminders to take medication every morning at a random time between 7 AM and 8 AM. (The reminders included phrases such as "Please remember to take today's medications" and "It is time to take today's medications. Do not stop taking your medications without consulting a cardiologist.") The time participants received the reminders was selected based on feedback that we collected in the pilot study. The educational materials and reminders were sent through Message Express on an encrypted external device (an iPhone 5 borrowed from Duke University that was used specifically for this study). The intervention lasted for 60 days for each participant. After the intervention, participants were observed for 30 days.

Figure 1. Brief description of the 2 apps.

Data Collection and Variables

Participants' demographic characteristics were recorded, including age, gender, ethnicity, weight, height, marital status, job status, education, medical insurance status, living area, number of prescribed medications, and family income. A REDCap survey hyperlink was sent to each participant to collect their demographic characteristics and track their health outcome variables at 7 time points: enrollment (ie, baseline); 15, 30, 45, and 60 days after enrollment (ie, at the end of the intervention); and 75 and 90 days after enrollment (ie, at the end of follow-up). Health outcome variables included medication nonadherence score, heart rate (HR), systolic blood pressure (SBP), and diastolic blood pressure (DBP). The primary outcome (ie, medication nonadherence score) was measured using a validated 3-item, 5-point Likert scale, the Voils Extent Scale [42]. To calculate the total nonadherence score, the responses to the relevant 3 items of the survey were added up. A lower score indicated better medication adherence. If a participant did not complete a survey, the study coordinator sent them a WeChat message as a reminder. If there was no reply after 2 such attempts, the study coordinator called the participant. If there was still no answer, the participant was treated as lost to follow-up. Participants who deleted the study coordinator's WeChat account were treated as having withdrawn from the study.

Data Analysis

All analyses were performed using SAS 9.4. Categorical variables such as gender, ethnicity, marital status, job status, medical insurance, and living area were reported as frequency

(n) and percentage (%). Numerical variables including age, weight, height, and number of prescribed medications were reported as mean (SD). The chi-square test and independent *t* tests were used to determine between-group differences at baseline. The Fisher exact test and Wilcoxon 2-sample test were used as alternatives when a chi-square test was not applicable due to small expected values in some cells or when a 2-sample *t* test assumption was not met. All statistical tests were nondirectional (2-tailed) and the significance level was set at $P=.05$. If a difference in a baseline characteristic was significant, the baseline variable was included as a covariate in the modeling stage.

The mean (SD) for all outcome variables (ie, medication nonadherence, HR, SBP, and DBP) in each group was computed and graphed to show trends at the 7 time points described above. Differences in outcomes between the experimental and control groups at the critical time points (ie, baseline, 60 days, and 90 days) were determined using the *t* test. Both the control and experimental groups contained individuals with extremely low or high blood pressure and HR, and calculating the mean of these abnormal values was likely to produce values that were misleadingly normal. Therefore, we dichotomized SBP, DBP, and HR values into binary variables: normal and abnormal. We defined normal blood pressure according to the 2010 Chinese Hypertension Guidelines [43], which define normal SBP and DBP as less than 140 mmHg and less than 90 mmHg, respectively, and define normal HR as between 60 bpm and 100 bpm. The proportional rates of normal HR, SBP, and DBP in the subjects were graphed to show trends. To test between-group differences in the trajectory of change in all outcomes after

adjusting for relevant covariates, a mixed-effects model was used for continuous outcomes and a generalized mixed-effects model with a logit link was used for the dichotomized binary outcomes. Considering the correlations between repeated longitudinal measurements of the same patient, a random intercept, a random slope, and an unstructured covariate matrix were used. The fixed effects included in the 2 models were the intervention group (ie, experimental versus control), time, the group-by-time interaction, and baseline variables that were significantly different in the experimental and control groups or covariates that could clinically impact HR and blood pressure. We also accommodated a quadratic time trend by including a time-by-time fixed effect and testing for it. If the quadratic effect was not significant, we removed it from the mixed-effects model.

Results

Baseline Characteristics

In this study, we recruited 230 participants and randomly assigned 116 to the experimental group and 114 to the control group (see Figure 2 for the CONSORT [Consolidated Standards of Reporting Trials] diagram [44]). Of the 230 participants, 34

participants did not complete the baseline questionnaire and thus did not receive the intervention. We collected baseline data from the 196 participants who received the intervention; of these, 6 participants later dropped out of the study and 9 were lost during the follow-up period. There was no statistically significant difference ($P>.05$) in any baseline characteristic between the 2 groups (Table 1). The majority of the participants were married (181/196, 92.4%), male (157/196, 80.1%), Han Chinese (184/196, 93.9%), and living in urban China (161/196, 82.1%). Participants' average age was 61 years, and half were retired (103/191, 53.9%). More than half the participants (121/196, 61.7%) were prescribed at least 5 medications. Half the participants had an annual family income less than CNY ¥54,000 (US \$7950). The majority of participants (186/192, 96.9%) had medical insurance; nearly 28% (41/147) of participants had their medications fully covered, 60% (89/147) were partially covered, and 11.6% (17/147) of participants did not have insurance coverage for their medication. Most participants (109/196, 88%) had used WeChat for more than a year and were using WeChat every day (152/188, 80.9%) before participating in the study; only 5.8% (11/188) had never used WeChat before the study. There was no difference between the experimental and control groups in mobile phone literacy.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flow diagram.

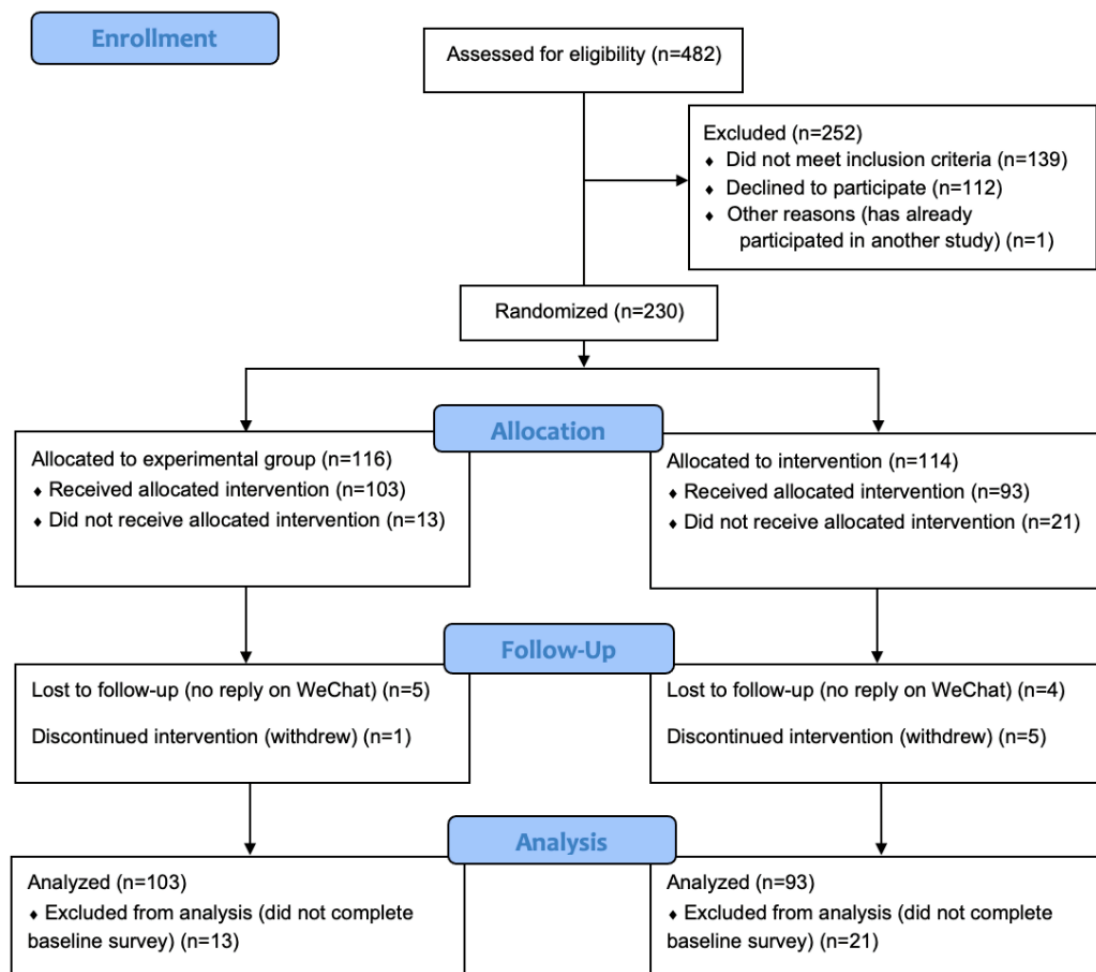


Table 1. Baseline characteristics of participants at enrollment.

Variable	All participants (N=196)	Experimental group (n=103)	Control group (n=93)	P value
Gender, n (%)				.86
Male	157 (80.1)	83 (80.6)	74 (79.6)	
Female	39 (19.9)	20 (19.4)	19 (20.4)	
Age (years), mean (SD)	61 (11)	61 (11)	62 (11)	.48
Height (cm), mean (SD)	164.8 (8.0)	165.5 (8.0)	164.1 (8.0)	.23
Weight (kg), mean (SD)	67.6 (11.3)	68.2 (11.9)	66.9 (10.5)	.45
Ethnicity, n (%)				.98 ^a
Han	184 (93.9)	96 (93.2)	88 (94.6)	
Zang	6 (3.1)	3 (2.9)	3 (3.2)	
Yi	1 (0.5)	1 (1.0)	0 (0)	
Hui	1 (0.5)	0 (0)	1 (1.1)	
Mongolian	1 (0.5)	1 (1.0)	0 (0)	
Other ethnic minorities	3 (1.5)	2 (1.9)	1 (1.1)	
Marital status, n (%)				.25
Married	181 (92.4)	93 (90.3)	88 (94.6)	
Widowed, separated, divorced, or single	15 (7.6)	10 (9.7)	5 (5.4)	
Job status, n (%)				.31 ^a
Employed	65 (34.0)	39 (39.0)	26 (28.6)	
Unemployed	4 (2.1)	3 (3.0)	1 (1.1)	
Farmer	19 (10.0)	8 (8.0)	11 (12.1)	
Retired	103 (53.9)	50 (50.0)	53 (58.2)	
Education, n (%)				.36
Primary school or lower	31 (16.1)	16 (15.8)	15 (16.3)	
Middle school	43 (22.3)	27 (26.7)	16 (17.4)	
High school	47 (24.3)	26 (25.7)	21 (22.8)	
Noncollege postsecondary	28 (14.5)	11 (10.9)	17 (18.5)	
College or above	44 (22.8)	21 (20.8)	23 (25.0)	
Number of prescribed medications, n (%)				.82 ^a
<5	75 (38.3)	41 (39.8)	34 (36.6)	
5-9	116 (59.2)	59 (57.3)	57 (61.3)	
≥10	5 (2.5)	3 (2.9)	2 (2.1)	
Covered by medical insurance	186 (96.9)	97 (96.0)	89 (97.8)	.69 ^a
Medication coverage as part of medical insurance, n (%)				.90 ^b
Complete coverage	41 (27.9)	22 (30.1)	19 (25.7)	
Some coverage	89 (60.5)	40 (54.8)	49 (66.2)	
No coverage	17 (11.6)	11 (15.1)	6 (8.1)	
Place of residence, n (%)				.82
Urban	161 (82.1)	84 (81.6)	77 (82.8)	
Rural	35 (17.9)	19 (18.4)	16 (17.2)	
Living arrangements, n (%)				.36
Living alone	14 (7.1)	9 (8.7)	5 (5.4)	

Variable	All participants (N=196)	Experimental group (n=103)	Control group (n=93)	P value
Living with family or relatives	182 (92.9)	94 (91.3)	88 (94.6)	
Length of using WeChat, n (%)				.78
<1 year	24 (12.0)	11 (10.7)	13 (14.0)	
1-4 years	109 (56.0)	58 (56.3)	51 (54.8)	
≥5 years	63 (32.0)	34 (33.0)	29 (31.2)	
Frequency of WeChat use before participating in the study, n (%)				.44
Daily	152 (80.9)	83 (83.8)	69 (77.5)	
Occasionally	25 (13.3)	12 (12.1)	13 (14.6)	
Never	11 (5.8)	4 (4.0)	7 (7.9)	
Annual family income (CNY¥), n (%)				.12 ^b
<¥54,000	81 (52.3)	48 (59.3)	33 (44.6)	
¥54,001-¥90,000	34 (21.9)	13 (16.1)	21 (28.4)	
¥90,001-¥120,000	18 (11.6)	11 (13.6)	7 (9.5)	
>¥120,000	22 (14.2)	9 (11.1)	13 (17.6)	
General health status, n (%)				.89
Good	70 (35.7)	37 (35.9)	33 (35.5)	
Fair	101 (51.5)	54 (52.4)	47 (50.5)	
Bad	25 (12.8)	12 (11.7)	13 (14.0)	

^aThe Fisher exact test was used due to small values in cells.

^bThe Mann-Whitney *U* test was used due to missing data: 25% of participants did not know whether their prescribed medications would be covered by their medical insurance at the time of discharge; and 21% of participants refused to disclose their family income.

Medication Nonadherence and Health Outcomes

The medication nonadherence score (Figure 3) consistently decreased in the experimental group. The nonadherence score decreased in the control group during the first 45 days, but then increased and approached the baseline level at 90 days (the score at 90 days was 0.08 less than baseline). The mean decrease in medication nonadherence score in the experimental group was greater than the mean decrease in the control group at 60 days ($t_{179}=2.04$, $P=.04$) and 90 days ($t_{155}=3.48$, $P<.001$) (Table 2).

HR decreased (Figure 4) at 60 days and 90 days in both groups compared to baseline, but the difference in the decrease between the 2 groups was not statistically significant at either 60 days ($t_{148}=-0.28$, $P=.78$) or 90 days ($t_{145}=0.32$, $P=.75$). SBP (Figure 5) and DBP (Figure 6) decreased at 60 days and 90 days in the experimental group but increased in the control group. The difference in the change in DBP between the intervention and control groups was statistically significant at both 60 days ($t_{160}=2.07$, $P=.04$) and 90 days ($t_{164}=2.21$, $P=.03$). The difference in the change in SBP between the 2 groups was statistically significant at 90 days ($t_{165}=3.12$, $P=.002$), but not significant at 60 days ($t_{161}=1.92$, $P=.06$).

The proportional rate of participants with normal SBP (Multimedia Appendix 1) and DBP (Multimedia Appendix 2) increased in both groups at 60 days and 90 days compared to baseline, but the difference between the 2 groups at these times was not statistically significant (Multimedia Appendix 3). Unlike

SBP and DBP, the proportional rate of participants with normal HR (Multimedia Appendix 4) decreased in both the experimental and control groups, but the difference between the 2 groups was not statistically significant at either 60 days ($P=.37$) or 90 days ($P=.41$).

Although no baseline characteristics were significantly different between the experimental group and the control group, HR and blood pressure can be influenced by body weight [45,46], gender [47,48], and age [49]; therefore, they were included as covariates in the modeling stage. After controlling for the effects of baseline body weight, gender, age, educational level, group, time, and the group-by-time interaction in the mixed-effects model, the difference in the rate of change of medication nonadherence between the 2 groups was statistically significant ($\beta=-.02$, $P<.001$) (Table 3).

The difference in the rate of change of SBP ($\beta=-.08$, $P<.001$) and DBP ($\beta=-.05$, $P=.004$) between the 2 groups was also statistically significant. However, the difference in the rate of change in HR between the 2 groups was not statistically significant ($\beta=-.01$, $P=.74$). In the generalized mixed effect model with a logit link for dichotomized binary outcomes, we found that the difference in the change of the proportional rate of participants with a normal SBP between the 2 groups was statistically significant ($\beta=.01$, $P=.02$), but that the difference between the 2 groups was not statistically significant for either the proportional rate of participants with normal DBP ($\beta=.006$, $P=.32$) or normal HR ($\beta=.003$, $P=.60$) (Multimedia Appendix 5).

Figure 3. Comparison of changes in medication nonadherence score between groups (SE 2).

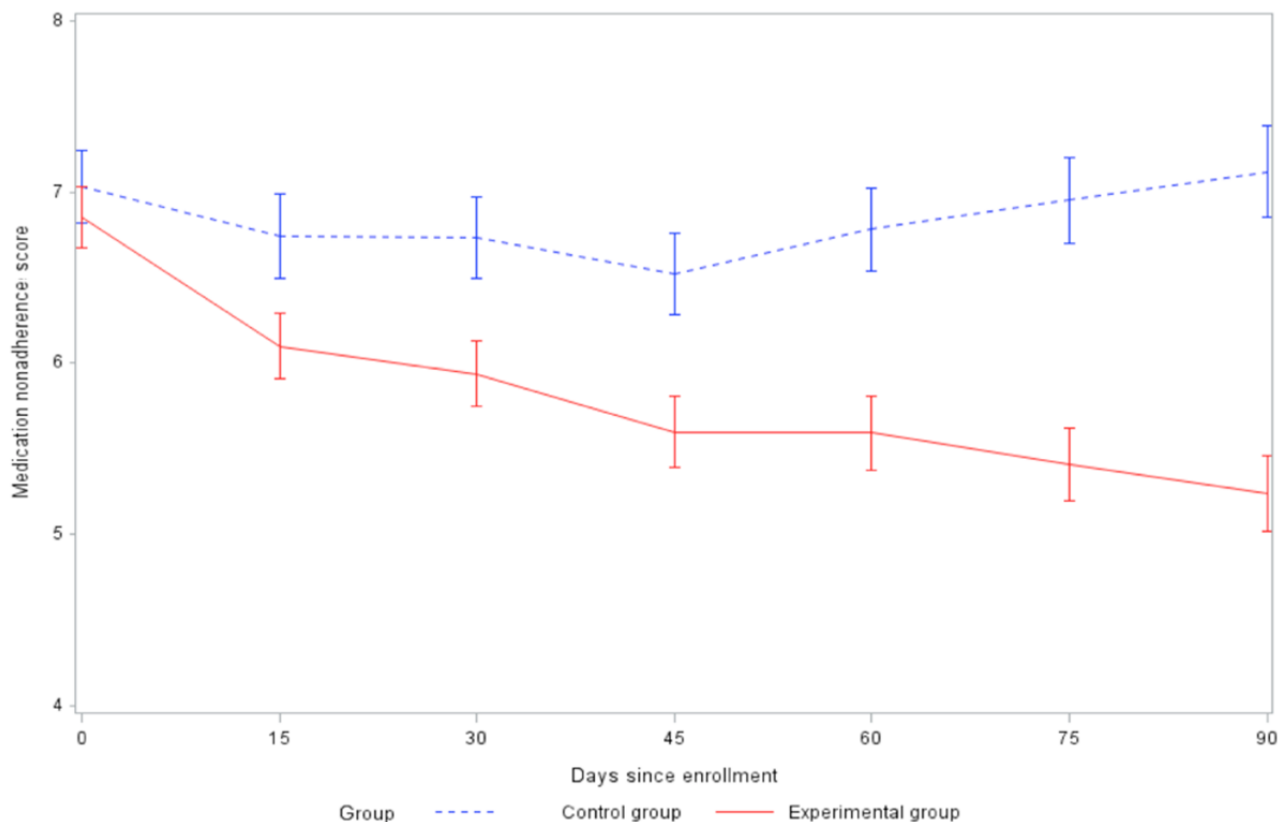


Table 2. Comparison of baseline and unadjusted changes in medication nonadherence and health outcomes between the 2 arms.

	Experimental group (n=103), mean (SD)	Control group (n=93), mean (SD)	t test (df)	P value
Medication nonadherence score				
Baseline	6.85 (1.85)	7.03 (2.05)	0.64 (194)	.52
Day 60	-1.21 (2.59)	-0.42 (2.63)	2.04 (179)	.04
Day 90	-1.58 (2.49)	-0.08 (3.15)	3.48 (155)	<.001
Heart rate (bpm)				
Baseline	73.93 (12.03)	73.16 (9.76)	-0.49 (183)	.63
Day 60	-1.46 (12.68)	-1.95 (9.03)	-0.28 (148)	.78
Day 90	-1.88 (12.88)	-1.32 (8.98)	0.32 (145)	.75
Systolic blood pressure (mmHg)				
Baseline	129.5 (14.37)	125.2 (15.10)	-2.04 (188)	.04
Day 60	-2.14 (16.20)	2.72 (16.07)	1.92 (161)	.06
Day 90	-2.87 (15.10)	4.38 (14.89)	3.12 (165)	.002
Diastolic blood pressure (mmHg)				
Baseline	78.71 (11.73)	76.03 (15.98)	-1.30 (162)	.20
Day 60	-2.46 (12.49)	1.87 (14.06)	2.07 (160)	.04
Day 90	-1.57 (12.21)	2.92 (13.99)	2.21 (164)	.03

Figure 4. Comparison of changes in heart rate between groups (SE 2).

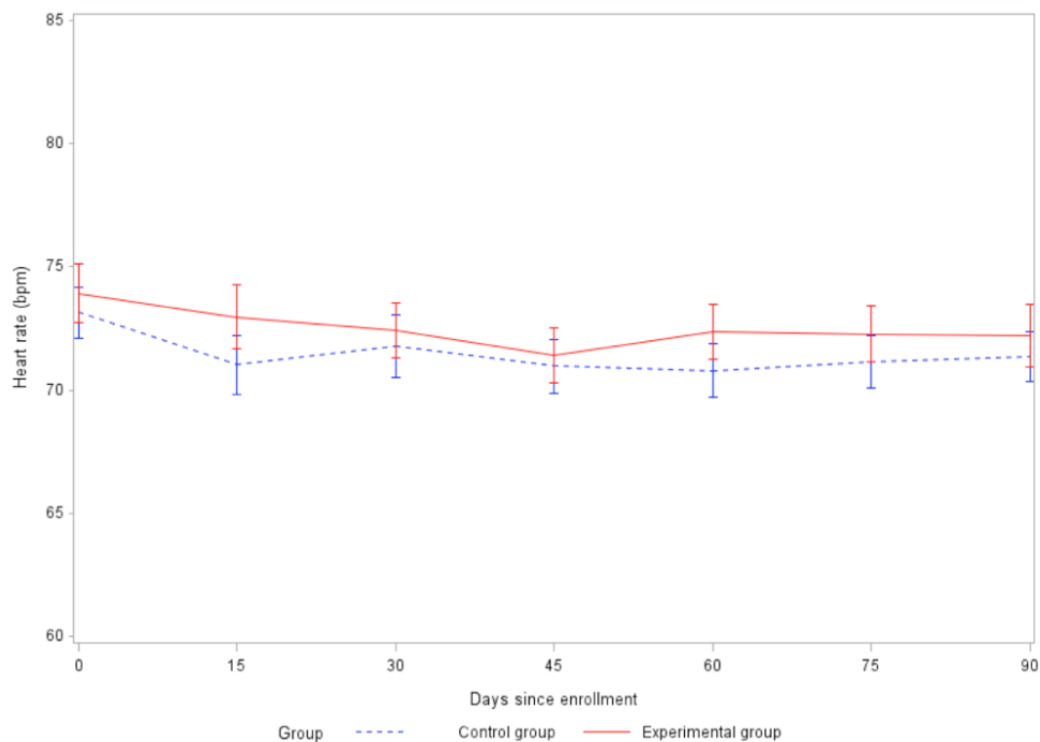


Figure 5. Comparison of changes in systolic blood pressure between groups (SE 2).

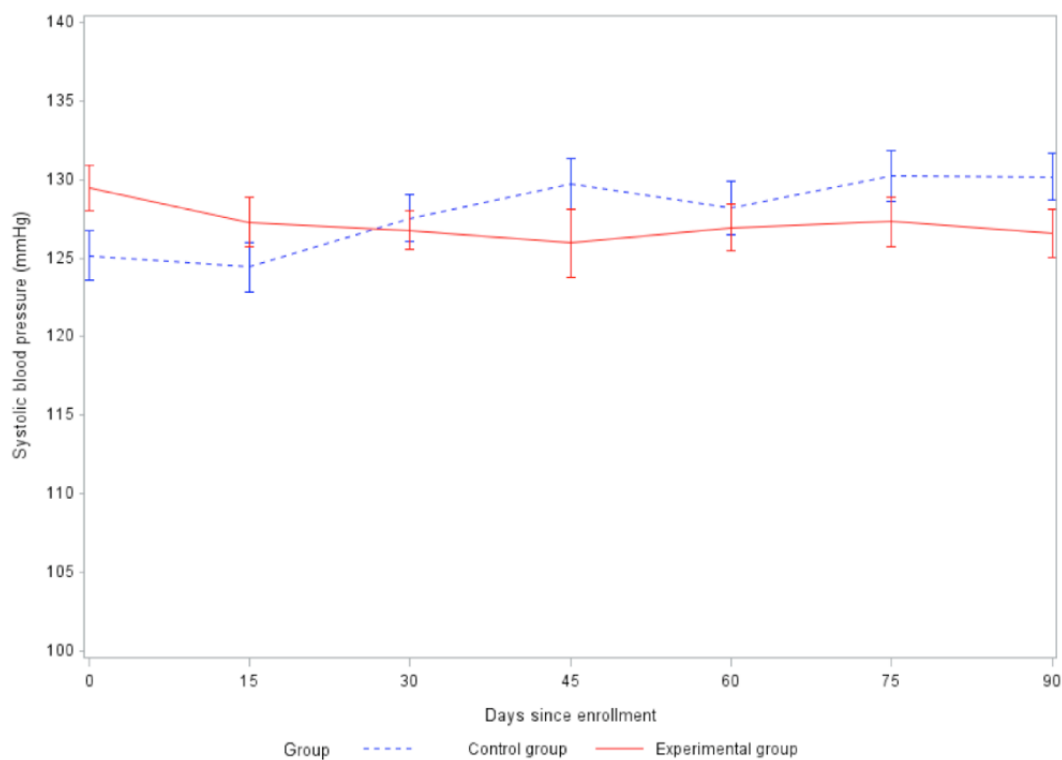


Figure 6. Comparison of changes in diastolic blood pressure between groups (SE 2).

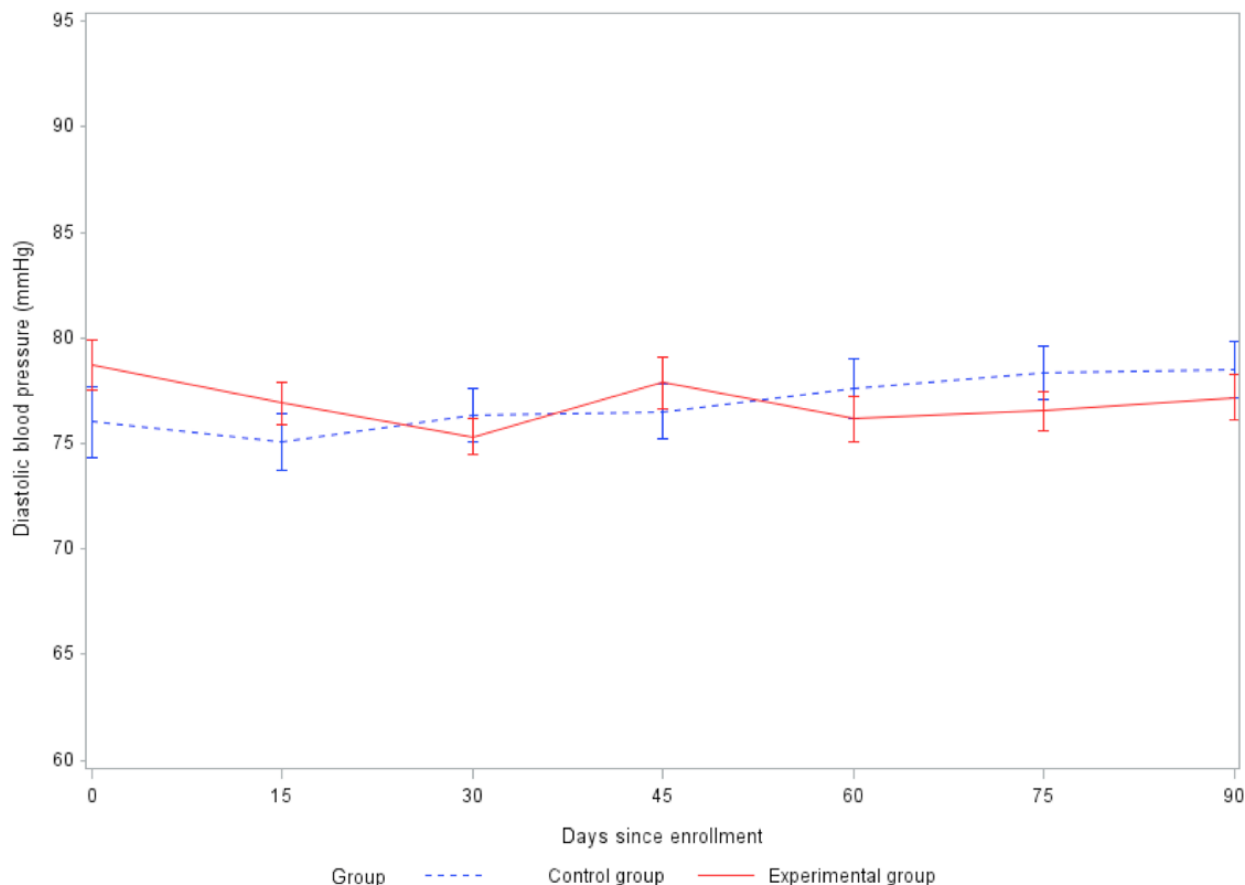


Table 3. Results of a mixed-effects model with a random intercept and slope to measure between-group differences in the trajectory of changes in outcomes.

Item	Medication nonadherence		Diastolic blood pressure (mmHg)		Systolic blood pressure (mmHg)		Heart rate (bpm)	
	Parameter estimation	P value	Parameter estimation	P value	Parameter estimation	P value	Parameter estimation	P value
Intercept	5.11	<.001	84.86	<.001	93.48	<.001	56.66	<.001
Group (reference: control)	-0.23	.37	2.05	.19	3.39	.06	0.68	.63
Gender (reference: male)	0.29	.33	-2.50	.15	0.23	.92	2.00	.26
Age	0.002	.85	-0.22	<.001	0.26	.001	0.001	.99
Education	0.11	.17	0.17	.72	0.41	.50	-0.21	.66
Weight	0.013	.27	0.11	.12	0.21	.02	0.21	.003
Time (days)	0.001	.74	0.04	.14	0.06	.09	-0.01	.33
Group × time	-0.02	<.001	-0.05	.004	-0.08	<.001	-0.01	.74

Discussion

Principal Findings

The aim of this study was to assess if a mobile phone-based mHealth intervention could improve medication adherence and relevant health outcomes (eg, blood pressure and HR) among patients with CHD in comparison to a control group that received general educational materials over a period of 2 months. We found that our mHealth intervention increased medication adherence and had a lasting effect in improving

medication adherence among patients with CHD, even though no intervention was given after 60 days. We also found that our mHealth intervention improved health outcomes by lowering SBP and DBP, and that this effect continued for 30 days after the intervention. Unlike SBP and DBP, our mHealth intervention did not significantly lower HR, although mean HR consistently remained within the normal range. After dichotomizing SBP, DBP, and HR into binary variables, we found that there was no significant difference between the proportional rate of

participants with normal SBP, DBP, and HR at baseline, 60 days, or 90 days between the intervention and control groups.

A mixed-effects model for continuous outcomes showed that differences in the rate of change of medication nonadherence, SBP, and DBP were statistically significant between the 2 groups after controlling for the effects of group, time, the group-by-time interaction, and some baseline variables that can influence HR and blood pressure, such as body weight, gender, age, and educational level. Similarly, the generalized mixed-effects model with a logit link used for dichotomized binary outcomes showed that after controlling for the effects of baseline body weight, gender, age, educational level, group, time, and the group-by-time interaction, our mHealth intervention could significantly increase the proportional rate of participants with normal SBP. In summary, these key findings demonstrate that our mHealth intervention could significantly increase medication adherence and the proportional rate of participants with normal SBP, as well as lower SBP and DBP, in patients with CHD.

In this study, participants were defined as dropping out if (1) they filled out the baseline survey but deleted the study coordinator's WeChat account at any time point before their completion of the study; or (2) they told the study coordinator that they wanted to withdraw from the study. Participants were defined as lost to follow-up if they filled out the baseline survey and kept the study coordinator's WeChat account, but did not fill out later surveys and did not respond to reminders and WeChat calls made by the study coordinator. This could happen at any time point during the study. In this study, 9 participants were lost to follow-up and 6 participants dropped out, which together accounts for 7% of the data. Due to the small amount of missing data, we did not perform missing data imputation. We compared the baseline variables between participants who dropped out or were lost to follow-up with those who completed the study, and found that none of the baseline variables were significantly different between these groups ([Multimedia Appendix 6](#)).

In this study, the ratio of male to female patients with CHD was 4:1, which is the same ratio as in our pilot study conducted in 2017. The higher proportion of male participants reflects the higher prevalence of CHD in men in China; Chinese men are more likely to engage in risky behaviors, such as smoking and drinking [12]. This result is consistent with findings that US men had a higher prevalence of CHD than women in all age strata from 45 to 94 years, according to the American Heart Association's report *Heart Disease and Stroke Statistics* [50,51].

Limitations

This study contributes important knowledge about mHealth as a tool to improve medication adherence, but it has several limitations. First, all participants were recruited at a major university-affiliated hospital; many of them were middle-aged male urban residents and were thus not representative of the general Chinese population. Second, health outcomes such as SBP, DBP, and HR were self-reported by participants. There may therefore have been discrepancies regarding these measurements. Third, during this study, some participants did not provide their health outcome data in a timely manner. To

collect these data, some participants had to be reminded by WeChat messages and phone calls. These messages and phone calls might have been covariates that could influence the participants' medication-taking behaviors, but we did not consider their influence. Finally, the mHealth intervention was not automated and lasted for only 60 days, a comparatively short period of time. To adopt this intervention in real clinical settings, which have thousands of patients, and to create long-term effects, it will be necessary to use automated methods. Finally, this study was unblinded. All participants, data collectors, and data analysts were aware of which treatment arms participants had been assigned to. To increase rigor and reduce bias, future large multisite studies should consider a double-blind design.

Future Directions

The treatment of many chronic illnesses involves long-term pharmaceutical therapy. Nevertheless, it is an ongoing challenge to find effective ways to improve medication adherence and other health behaviors to improve health outcomes. For patients with CHD, cardioprotective medications can prevent the enlargement of harmful clots [52], cardiovascular symptoms, and poor therapeutic outcomes. Poor adherence to cardioprotective medications, however, is considered to be a public health concern [18,52]. It has been linked to increases in health care costs due to poor therapeutic outcomes that typically require major medical interventions, such as coronary angioplasty and coronary artery bypass grafting [20]. Mobile phones are widely used in China; 97% of Chinese netizens access the internet through a mobile phone [30]. This makes mobile phones an ideal platform for implementing an mHealth intervention to improve medication adherence. If this mHealth intervention could be scaled up for use in real-world clinical settings, it would have 3 useful aspects. First, it is able to overcome geographic barriers. This is particularly useful in China, considering the fact that the majority of its highly qualified health care providers reside in cities, and most of its tertiary hospitals are located in urban areas [53,54]. Hospitals in China are categorized into a 3-tier system: primary, secondary, and tertiary [55], with tertiary being the highest level of quality. To seek quality medical care, many rural residents have to travel to urban areas to find qualified health care providers. This geographic barrier limits the access of rural patients with CHD to quality medical care and may impact their health outcomes across their lifespans. Our mHealth intervention has the potential to increase the access of patients with CHD to quality medical care through mobile phones without requiring them to travel to hospitals. Second, the core technology of this mHealth intervention is app-based mobile messaging, which is widely used in China. Therefore, this intervention could be a cost-effective method to remind people across socioeconomic strata and across geographic areas to take their medications. If this intervention were automated, and if it resulted in even modest improvements in medication adherence and blood pressure, there could be population-level benefits. Given the economic burden and prevalence of CHD in China, even small clinical improvements could have a large impact. In other words, the cost-benefit ratio of this mHealth intervention is expected to be positive.

Although some studies have been conducted in China on using mobile apps (eg, WeChat) to improve blood pressure and self-management behavior [31-34], none have been designed specifically to address the problem of medication nonadherence in patients with CHD. Previous interventions have also not been individualized. For example, in an mHealth intervention study conducted in Guangzhou, educational materials were sent to a group of participants rather than to individuals [34]. Moreover, no previous study has used an mHealth intervention that integrated 2 mobile apps. Given that medication nonadherence is a complex issue involving multiple factors related to patients, health care providers, and health care systems [21], integrating mobile apps with different functionalities might be a promising approach. The mHealth intervention that we tested in this study could be harnessed and adapted to other fields, such as endocrinology, orthopedics, and mental health. Finally, adopting this intervention is integral to developing a platform for immediate medical consultation and emergency health care. Cardiovascular events such as heart attack, heart failure,

arrhythmia, and heart valve problems [1] are emergencies in which many patients need immediate medical treatment, consultation, and guidance. Adopting this mHealth intervention is an important step for hospitals to develop online tools for immediate health care service.

Conclusion

The treatment of many chronic illnesses involves long-term pharmaceutical therapy, but it is an ongoing challenge to find effective ways to improve medication adherence and promote good health outcomes. In this study, we examined an mHealth intervention to remind patients with CHD to take their cardioprotective medications. Our results demonstrate that it is feasible to conduct an mHealth intervention to improve medication adherence and health outcomes, represented by measures including SBP and DBP. In summary, mHealth interventions that are constructed using evidence-based content show promise to help increase medication adherence and improve health outcomes.

Acknowledgments

This study was supported by the Duke University Global Health Institute (2018 Duke Global Health Doctoral Certificate Fieldwork Grant); Duke University Graduate School (2017 International Dissertation Research Travel Award); and the Duke University School of Nursing (2018 PhD Student Pilot Study Fund).

Authors' Contributions

ZN contributed to conceptualization, methodology, investigation, and formal analysis and wrote the original draft. BW contributed to conceptualization, methodology, validation, supervision, writing, reviewing, and editing. QY contributed to conceptualization, methodology, validation, formal analysis, writing, reviewing, and editing. LLY contributed to conceptualization, methodology, validation, writing, reviewing, and editing. CL contributed to methodology, project coordination, writing, reviewing, and editing. RJS contributed to conceptualization, methodology, validation, supervision, writing, reviewing, and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Comparison of changes of the proportional rates of normal SBP.

[PNG File , 103 KB - [jmir_v24i3e27202_app1.png](#)]

Multimedia Appendix 2

Comparison of changes of the proportional rates of normal DBP.

[PNG File , 104 KB - [jmir_v24i3e27202_app2.png](#)]

Multimedia Appendix 3

Frequency and proportion of normal blood pressures at multiple time points [n (%)].

[PDF File (Adobe PDF File), 19 KB - [jmir_v24i3e27202_app3.pdf](#)]

Multimedia Appendix 4

Comparison of changes of the proportional rates of normal heart rate.

[PNG File , 101 KB - [jmir_v24i3e27202_app4.png](#)]

Multimedia Appendix 5

Results of the mixed-effects model for proportional rate change of normal health outcomes.

[PDF File (Adobe PDF File), 27 KB - [jmir_v24i3e27202_app5.pdf](#)]

Multimedia Appendix 6

Baseline characteristics of participants who completed the study and those who did not [n (%)].

[PDF File (Adobe PDF File), 45 KB - [jmir_v24i3e27202_app6.pdf](#)]

Multimedia Appendix 7

CONSORT-EHEALTH (V 1.6.1) checklist.

[PDF File (Adobe PDF File), 324 KB - [jmir_v24i3e27202_app7.pdf](#)]

References

1. What is Cardiovascular Disease? American Heart Association. 2017 May 31. URL: <https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease> [accessed 2019-02-20]
2. The top 10 causes of death. World Health Organization. 2020 Dec 09. URL: <https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death> [accessed 2021-02-01]
3. Wilson PWF, Douglas PS. Epidemiology of coronary heart disease. UpToDate. 2017. URL: https://www.uptodate.com/contents/epidemiology-of-coronary-heart-disease?source=see_link [accessed 2022-03-04]
4. Bowry ADK, Lewey J, Dugani SB, Choudhry NK. The Burden of Cardiovascular Disease in Low- and Middle-Income Countries: Epidemiology and Management. *Can J Cardiol* 2015 Sep;31(9):1151-1159. [doi: [10.1016/j.cjca.2015.06.028](https://doi.org/10.1016/j.cjca.2015.06.028)] [Medline: [26321437](#)]
5. Gaziano TA, Bitton A, Anand S, Abrahams-Gessel S, Murphy A. Growing epidemic of coronary heart disease in low- and middle-income countries. *Curr Probl Cardiol* 2010 Feb;35(2):72-115 [FREE Full text] [doi: [10.1016/j.cpcardiol.2009.10.002](https://doi.org/10.1016/j.cpcardiol.2009.10.002)] [Medline: [20109979](#)]
6. The Global Burden of Disease. World Health Organization. 2004. URL: http://www.who.int/healthinfo/global_burden_disease/GBD_report_2004update_full.pdf [accessed 2019-11-20]
7. China remains largest developing country: Economist. China Daily. 2018. URL: <http://www.chinadaily.com.cn/a/201804/16/WS5ad4343fa3105cdcf65187d6.html> [accessed 2019-02-10]
8. China Overview. The World Bank. URL: <http://www.worldbank.org/en/country/china/overview> [accessed 2019-02-10]
9. China: WHO statistical profile. World Health Organization. 2015. URL: <https://www.who.int/gho/countries/chn.pdf?ua=1> [accessed 2019-01-22]
10. Zhang X, Lu ZL, Liu L. Coronary heart disease in China. *Heart* 2008 Sep;94(9):1126-1131. [doi: [10.1136/hrt.2007.132423](https://doi.org/10.1136/hrt.2007.132423)] [Medline: [18703693](#)]
11. Le C, Fang Y, Linxiong W, Shulan Z, Golden AR. Economic burden and cost determinants of coronary heart disease in rural southwest China: a multilevel analysis. *Public Health* 2015 Jan;129(1):68-73. [doi: [10.1016/j.puhe.2014.11.002](https://doi.org/10.1016/j.puhe.2014.11.002)] [Medline: [25542739](#)]
12. Chen W, Gao R, Liu L, Zhu M, Wang W, Wang Y, et al. China cardiovascular diseases report 2015: a summary. *J Geriatr Cardiol* 2017 Jan;14(1):1-10 [FREE Full text] [doi: [10.11909/j.issn.1671-5411.2017.01.012](https://doi.org/10.11909/j.issn.1671-5411.2017.01.012)] [Medline: [28270835](#)]
13. Coronary Heart Disease - Treatment. National Heart, Lung, and Blood Institute. URL: <https://www.nhlbi.nih.gov/health-topics/ischemic-heart-disease#Medicines> [accessed 2019-02-01]
14. Adherence to long-term therapies. World Health Organization. 2003. URL: https://www.who.int/chp/knowledge/publications/adherence_full_report.pdf?ua=1ISBN:978-92-4-154599-0 [accessed 2020-02-26]
15. Hamm CW, Bassand JP, Agewall S, Bax J, Boersma E, Bueno H, ESC Committee for Practice Guidelines. ESC Guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: The Task Force for the management of acute coronary syndromes (ACS) in patients presenting without persistent ST-segment elevation of the European Society of Cardiology (ESC). *Eur Heart J* 2011 Dec;32(23):2999-3054. [doi: [10.1093/eurheartj/ehr236](https://doi.org/10.1093/eurheartj/ehr236)] [Medline: [21873419](#)]
16. Zhang H, Yuan X, Zhang H, Chen S, Zhao Y, Hua K, et al. Efficacy of Long-Term β -Blocker Therapy for Secondary Prevention of Long-Term Outcomes After Coronary Artery Bypass Grafting Surgery. *Circulation* 2015 Jun 23;131(25):2194-2201 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.114.014209](https://doi.org/10.1161/CIRCULATIONAHA.114.014209)] [Medline: [25908770](#)]
17. Ho PM, Bryson CL, Rumsfeld JS. Medication adherence: its importance in cardiovascular outcomes. *Circulation* 2009 Jun 16;119(23):3028-3035. [doi: [10.1161/CIRCULATIONAHA.108.768986](https://doi.org/10.1161/CIRCULATIONAHA.108.768986)] [Medline: [19528344](#)]
18. Jiang J, Hong T, Yu R, Zhang Y, Liu Z, Huo Y. Knowledge of secondary prevention guidelines for coronary heart disease: results from a physicians' survey in China. *Eur J Prev Cardiol* 2012 Oct;19(5):991-998. [doi: [10.1177/1741826711421299](https://doi.org/10.1177/1741826711421299)] [Medline: [21859776](#)]
19. 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. *International Journal of Nursing Sciences* 2018 Oct;5(4):322-330. [doi: [10.1016/j.ijnss.2018.09.003](https://doi.org/10.1016/j.ijnss.2018.09.003)] [Medline: [31406843](#)]
20. Iuga AO, McGuire MJ. Adherence and health care costs. *Risk Manag Healthc Policy* 2014;7:35-44 [FREE Full text] [doi: [10.2147/RMHP.S19801](https://doi.org/10.2147/RMHP.S19801)] [Medline: [24591853](#)]

21. Ni Z, Dardas L, Wu B, Shaw R. Cardioprotective medication adherence among patients with coronary heart disease in China: a systematic review. *Heart Asia* 2019;11(2):e011173 [FREE Full text] [doi: [10.1136/heartasia-2018-011173](https://doi.org/10.1136/heartasia-2018-011173)] [Medline: [31297162](https://pubmed.ncbi.nlm.nih.gov/31297162/)]
22. Ding R, Ma C, Chen H, Wu Y, Yang X, Hua Q, et al. [Control rate of increased low-density lipoprotein cholesterol levels in cardiology outpatients with coronary heart disease in Beijing]. *Zhonghua Xin Xue Guan Bing Za Zhi* 2013 Mar;41(3):251-255. [Medline: [23879953](https://pubmed.ncbi.nlm.nih.gov/23879953/)]
23. Zhao S, Zhao H, Wang L, Du S, Qin Y. Education is critical for medication adherence in patients with coronary heart disease. *Acta Cardiol* 2015 Apr;70(2):197-204. [doi: [10.1080/ac.70.2.3073511](https://doi.org/10.1080/ac.70.2.3073511)] [Medline: [26148380](https://pubmed.ncbi.nlm.nih.gov/26148380/)]
24. Wang H, Zhang D, Hou Z, Yan F, Hou Z. Association between social health insurance and choice of hospitals among internal migrants in China: a national cross-sectional study. *BMJ Open* 2018 Feb 10;8(2):e018440 [FREE Full text] [doi: [10.1136/bmjopen-2017-018440](https://doi.org/10.1136/bmjopen-2017-018440)] [Medline: [29440156](https://pubmed.ncbi.nlm.nih.gov/29440156/)]
25. Meng Q, Yang H, Chen W, Sun Q, Liu X. People's Republic of China Health System Review. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/208229/9789290617280_eng.pdf [accessed 2019-11-18]
26. 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](https://pubmed.ncbi.nlm.nih.gov/24883008/)]
27. Boulos MNK, Brewer AC, Karimkhani C, Buller DB, Dellavalle RP. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online J Public Health Inform* 2014;5(3):229 [FREE Full text] [doi: [10.5210/ojphi.v5i3.4814](https://doi.org/10.5210/ojphi.v5i3.4814)] [Medline: [24683442](https://pubmed.ncbi.nlm.nih.gov/24683442/)]
28. mHealth for development: the opportunity of mobile technology for healthcare in the developing world. Vital Wave Consulting. URL: http://www.globalproblems-globalsolutions-files.org/unf_website/assets/publications/technology/mhealth/mHealth_for_Development_full.pdf [accessed 2019-11-20]
29. He Y. China's mobile users hit 1.3 billion in 2015.: *China Daily*; 2015. URL: http://www.chinadaily.com.cn/business/2016-01/26/content_23261207.htm [accessed 2019-01-22]
30. Statistical Report on Internet Development in China. China Internet Network Information Center. 2018. URL: <https://cnnic.com.cn/IDR/ReportDownloads/201807/P020180711391069195909.pdf> [accessed 2019-01-22]
31. Dorje T, Zhao G, Scheer A, Tsokey L, Wang J, Chen Y, et al. SMARTphone and social media-based Cardiac Rehabilitation and Secondary Prevention (SMART-CR/SP) for patients with coronary heart disease in China: a randomised controlled trial protocol. *BMJ Open* 2018 Jun 30;8(6):e021908 [FREE Full text] [doi: [10.1136/bmjopen-2018-021908](https://doi.org/10.1136/bmjopen-2018-021908)] [Medline: [29961032](https://pubmed.ncbi.nlm.nih.gov/29961032/)]
32. Dorje T, Zhao G, Tso K, Wang J, Chen Y, Tsokey L, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. *Lancet Digit Health* 2019 Nov;1(7):e363-e374 [FREE Full text] [doi: [10.1016/S2589-7500\(19\)30151-7](https://doi.org/10.1016/S2589-7500(19)30151-7)] [Medline: [33323210](https://pubmed.ncbi.nlm.nih.gov/33323210/)]
33. Li T, Ding W, Li X, Lin A. Mobile health technology (WeChat) for the hierarchical management of community hypertension: protocol for a cluster randomized controlled trial. *Patient Prefer Adherence* 2019;13:1339-1352 [FREE Full text] [doi: [10.2147/PPA.S215719](https://doi.org/10.2147/PPA.S215719)] [Medline: [31496664](https://pubmed.ncbi.nlm.nih.gov/31496664/)]
34. Li X, Li T, Chen J, Xie Y, An X, Lv Y, et al. A WeChat-Based Self-Management Intervention for Community Middle-Aged and Elderly Adults with Hypertension in Guangzhou, China: A Cluster-Randomized Controlled Trial. *Int J Environ Res Public Health* 2019 Oct 23;16(21):40-58 [FREE Full text] [doi: [10.3390/ijerph16214058](https://doi.org/10.3390/ijerph16214058)] [Medline: [31652688](https://pubmed.ncbi.nlm.nih.gov/31652688/)]
35. Park LG, Beatty A, Stafford Z, Whooley MA. Mobile Phone Interventions for the Secondary Prevention of Cardiovascular Disease. *Prog Cardiovasc Dis* 2016;58(6):639-650. [doi: [10.1016/j.pcad.2016.03.002](https://doi.org/10.1016/j.pcad.2016.03.002)] [Medline: [27001245](https://pubmed.ncbi.nlm.nih.gov/27001245/)]
36. Li X, Xu Z, Tang N, Ye C, Zhu X, Zhou T, et al. Effect of intervention using a messaging app on compliance and duration of treatment in orthodontic patients. *Clin Oral Investig* 2016 Nov;20(8):1849-1859. [doi: [10.1007/s00784-015-1662-6](https://doi.org/10.1007/s00784-015-1662-6)] [Medline: [26631059](https://pubmed.ncbi.nlm.nih.gov/26631059/)]
37. Luo L, Luo L, Zhang X, He X. Hospital daily outpatient visits forecasting using a combinatorial model based on ARIMA and SES models. *BMC Health Serv Res* 2017 Jul 10;17(1):469 [FREE Full text] [doi: [10.1186/s12913-017-2407-9](https://doi.org/10.1186/s12913-017-2407-9)] [Medline: [28693579](https://pubmed.ncbi.nlm.nih.gov/28693579/)]
38. Jacob C. Statistical power analysis for the behavioural sciences (second edition). Lawrence Earlbaum Associates: Hillsdale, NJ; 1988.
39. Cunningham JB, Gardner E. Power, effect and sample size using GPower: practical issues for researchers and members of research ethics committees. *Evidence Based Midwifery* 2007;5(4):132-136.
40. World Health Organization. 事实档案 [Facts in pictures]. URL: <https://www.who.int/zh/news-room/facts-in-pictures> [accessed 2017-01-01]
41. World Health Organization. 心血管疾病 [Cardiovascular diseases]. URL: https://www.who.int/zh/health-topics/cardiovascular-diseases#tab=tab_1 [accessed 2017-01-01]
42. Voils CI, Maciejewski ML, Hoyle RH, Reeve BB, Gallagher P, Bryson CL, et al. Initial validation of a self-report measure of the extent of and reasons for medication nonadherence. *Med Care* 2012 Dec;50(12):1013-1019 [FREE Full text] [doi: [10.1097/MLR.0b013e318269e121](https://doi.org/10.1097/MLR.0b013e318269e121)] [Medline: [22922431](https://pubmed.ncbi.nlm.nih.gov/22922431/)]
43. Liu L, Writing Group of 2010 Chinese Guidelines for the Management of Hypertension. [2010 Chinese guidelines for the management of hypertension]. *Zhonghua Xin Xue Guan Bing Za Zhi* 2011 Jul;39(7):579-615. [Medline: [22088239](https://pubmed.ncbi.nlm.nih.gov/22088239/)]

44. Schulz KF, Altman DG, Moher D, Fergusson D. CONSORT 2010 changes and testing blindness in RCTs. *Lancet* 2010 Apr 03;375(9721):1144-1146. [doi: [10.1016/S0140-6736\(10\)60413-8](https://doi.org/10.1016/S0140-6736(10)60413-8)] [Medline: [20338625](https://pubmed.ncbi.nlm.nih.gov/20338625/)]
45. Harsha DW, Bray GA. Weight loss and blood pressure control (Pro). *Hypertension* 2008 Jun;51(6):1420-1425. [doi: [10.1161/HYPERTENSIONAHA.107.094011](https://doi.org/10.1161/HYPERTENSIONAHA.107.094011)] [Medline: [18474829](https://pubmed.ncbi.nlm.nih.gov/18474829/)]
46. Seimon RV, Espinoza D, Finer N, James WPT, Legler UF, Coutinho W, et al. Changes in body weight and pulse: outcome events in overweight and obese subjects with cardiovascular disease in the SCOUT trial. *Int J Obes (Lond)* 2015 May;39(5):849-857. [doi: [10.1038/ijo.2014.211](https://doi.org/10.1038/ijo.2014.211)] [Medline: [25520249](https://pubmed.ncbi.nlm.nih.gov/25520249/)]
47. Maranon R, Reckelhoff JF. Sex and gender differences in control of blood pressure. *Clin Sci (Lond)* 2013 Oct;125(7):311-318 [FREE Full text] [doi: [10.1042/CS20130140](https://doi.org/10.1042/CS20130140)] [Medline: [23746374](https://pubmed.ncbi.nlm.nih.gov/23746374/)]
48. Reckelhoff JF. Gender differences in the regulation of blood pressure. *Hypertension* 2001 May;37(5):1199-1208. [doi: [10.1161/01.hyp.37.5.1199](https://doi.org/10.1161/01.hyp.37.5.1199)] [Medline: [11358929](https://pubmed.ncbi.nlm.nih.gov/11358929/)]
49. Buford TW. Hypertension and aging. *Ageing Res Rev* 2016 Mar;26:96-111 [FREE Full text] [doi: [10.1016/j.arr.2016.01.007](https://doi.org/10.1016/j.arr.2016.01.007)] [Medline: [26835847](https://pubmed.ncbi.nlm.nih.gov/26835847/)]
50. Mosca L, Barrett-Connor E, Wenger NK. Sex/gender differences in cardiovascular disease prevention: what a difference a decade makes. *Circulation* 2011 Nov 08;124(19):2145-2154 [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.110.968792](https://doi.org/10.1161/CIRCULATIONAHA.110.968792)] [Medline: [22064958](https://pubmed.ncbi.nlm.nih.gov/22064958/)]
51. Roger VL, Go AS, Lloyd-Jones DM, Adams RJ, Berry JD, Brown TM, American Heart Association Statistics Committee Stroke Statistics Subcommittee. Heart disease and stroke statistics--2011 update: a report from the American Heart Association. *Circulation* 2011 Feb 01;123(4):e18-e209 [FREE Full text] [doi: [10.1161/CIR.0b013e3182009701](https://doi.org/10.1161/CIR.0b013e3182009701)] [Medline: [21160056](https://pubmed.ncbi.nlm.nih.gov/21160056/)]
52. Bi Y, Gao R, Patel A, Su S, Gao W, Hu D, CPACS Investigators. Evidence-based medication use among Chinese patients with acute coronary syndromes at the time of hospital discharge and 1 year after hospitalization: results from the Clinical Pathways for Acute Coronary Syndromes in China (CPACS) study. *Am Heart J* 2009 Mar;157(3):509-516.e1. [doi: [10.1016/j.ahj.2008.09.026](https://doi.org/10.1016/j.ahj.2008.09.026)] [Medline: [19249422](https://pubmed.ncbi.nlm.nih.gov/19249422/)]
53. Xie X, Liu P, Zheng Y, Zhou W, Zou J, Wang X, et al. Equity of health resource distribution in China during 2009–15: an analysis of cross-sectional nationwide data. *The Lancet* 2017 Dec;390:S6. [doi: [10.1016/s0140-6736\(17\)33144-6](https://doi.org/10.1016/s0140-6736(17)33144-6)]
54. Wang H, Xu T, Xu J. Factors contributing to high costs and inequality in China's health care system. *JAMA* 2007 Oct 24;298(16):1928-1930. [doi: [10.1001/jama.298.16.1928](https://doi.org/10.1001/jama.298.16.1928)] [Medline: [17954544](https://pubmed.ncbi.nlm.nih.gov/17954544/)]
55. Li X, Huang J, Zhang H. An analysis of hospital preparedness capacity for public health emergency in four regions of China: Beijing, Shandong, Guangxi, and Hainan. *BMC Public Health* 2008 Sep 20;8:319 [FREE Full text] [doi: [10.1186/1471-2458-8-319](https://doi.org/10.1186/1471-2458-8-319)] [Medline: [18803860](https://pubmed.ncbi.nlm.nih.gov/18803860/)]

Abbreviations

CHD: coronary heart disease
DBP: diastolic blood pressure
HR: heart rate
mHealth: mobile health
SBP: systolic blood pressure

Edited by T Leung; submitted 18.01.21; peer-reviewed by P Kardas, N Cornelis, J Zhang; comments to author 30.06.21; revised version received 01.08.21; accepted 08.02.22; published 09.03.22.

Please cite as:

Ni Z, Wu B, Yang Q, Yan LL, Liu C, Shaw RJ

An mHealth Intervention to Improve Medication Adherence and Health Outcomes Among Patients With Coronary Heart Disease: Randomized Controlled Trial

J Med Internet Res 2022;24(3):e27202

URL: <https://www.jmir.org/2022/3/e27202>

doi: [10.2196/27202](https://doi.org/10.2196/27202)

PMID: [35262490](https://pubmed.ncbi.nlm.nih.gov/35262490/)

©Zhao Ni, Bei Wu, Qing Yang, Lijing L Yan, Changqing Liu, Ryan J Shaw. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 09.03.2022. 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.

Original Paper

An Innovative Telemedical Network to Improve Infectious Disease Management in Critically Ill Patients and Outpatients (TELnet@NRW): Stepped-Wedge Cluster Randomized Controlled Trial

Gernot Marx¹, Prof Dr; Wolfgang Greiner², Prof Dr; Christian Juhra³, PD Dr; Svenja Elkenkamp², MSc; Daniel Gensorowsky², MSc; Sebastian Lemmen⁴, Prof Dr; Jan Englbrecht³, Dr med; Sandra Dohmen¹, Dr med; Antje Gottschalk³, Prof Dr; Miriam Haverkamp⁴, Dr med; Annette Hempen⁵, MHBA; Christian Flügel-Bleienheuft⁶, Dr med; Daniela Bause³, MA; Henna Schulze-Steinen¹, MD; Susanne Rademacher¹, MSc; Jennifer Kistermann¹, MSc; Stefan Hoch⁶, Dipl.-Kfm.; Hans-Juergen Beckmann⁵, Dr med; Christian Lanckohr³, Dr med; Volker Lowitsch⁷, Dipl.-Math.; Arne Peine¹, Dr med; Fabian Juzek-Kuepper⁴, Dr med; Carina Benstoem¹, Prof Dr; Kathrin Sperling^{3*}, Dr med; Robert Deisz^{1*}, Dr med

¹Department of Intensive Care Medicine and Intermediate Care, Medical Faculty RWTH Aachen, Aachen, Germany

²School of Public Health, Department of Health Economics and Health Care Management, Bielefeld University, Bielefeld, Germany

³University Hospital Muenster, Muenster, Germany

⁴Division of Infection Control and Infectious Diseases, Medical Faculty RWTH Aachen, Aachen, Germany

⁵Physician Network, Medizin und Mehr eG (MuM), Buende, Germany

⁶Physician Network, Gesundheitsnetz Köln-Süd (GKS) e.V., Cologne, Germany

⁷Healthcare IT Solutions GmbH, Aachen, Germany

*these authors contributed equally

Corresponding Author:

Gernot Marx, Prof Dr

Department of Intensive Care Medicine and Intermediate Care, Medical Faculty RWTH Aachen

Pauwelsstr. 30

Aachen, 52074

Germany

Phone: 49 2418080444

Email: gmarx@ukaachen.de

Abstract

Background: Evidence-based infectious disease and intensive care management is more relevant than ever. Medical expertise in the two disciplines is often geographically limited to university institutions. In addition, the interconnection between inpatient and outpatient care is often insufficient (eg, no shared electronic health record and no digital transfer of patient findings).

Objective: This study aims to establish and evaluate a telemedical inpatient-outpatient network based on expert teleconsultations to increase treatment quality in intensive care medicine and infectious diseases.

Methods: We performed a multicenter, stepped-wedge cluster randomized trial (February 2017 to January 2020) to establish a telemedicine inpatient-outpatient network among university hospitals, hospitals, and outpatient physicians in North Rhine-Westphalia, Germany. Patients aged ≥ 18 years in the intensive care unit or consulting with a physician in the outpatient setting were eligible. We provided expert knowledge from intensivists and infectious disease specialists through advanced training courses and expert teleconsultations with 24/7/365 availability on demand respectively once per week to enhance treatment quality. The primary outcome was adherence to the 10 *Choosing Wisely* recommendations for infectious disease management. Guideline adherence was analyzed using binary logistic regression models.

Results: Overall, 159,424 patients (10,585 inpatients and 148,839 outpatients) from 17 hospitals and 103 outpatient physicians were included. There was a significant increase in guideline adherence in the management of *Staphylococcus aureus* infections (odds ratio [OR] 4.00, 95% CI 1.83-9.20; $P < .001$) and in sepsis management in critically ill patients (OR 6.82, 95% CI 1.27-56.61;

$P=.04$). There was a statistically nonsignificant decrease in sepsis-related mortality from 29% (19/66) in the control group to 23.8% (50/210) in the intervention group. Furthermore, the extension of treatment with prophylactic antibiotics after surgery was significantly less likely (OR 9.37, 95% CI 1.52-111.47; $P=.04$). Patients treated by outpatient physicians, who were regularly participating in expert teleconsultations, were also more likely to be treated according to guideline recommendations regarding antibiotic therapy for uncomplicated upper respiratory tract infections (OR 1.34, 95% CI 1.16-1.56; $P<.001$) and asymptomatic bacteriuria (OR 9.31, 95% CI 3.79-25.94; $P<.001$). For the other recommendations, we found no significant effects, or we had too few observations to generate models. The key limitations of our study include selection effects due to the applied on-site triage of patients as well as the limited possibilities to control for secular effects.

Conclusions: Telemedicine facilitates a direct round-the-clock interaction over broad distances between intensivists or infectious disease experts and physicians who care for patients in hospitals without ready access to these experts. Expert teleconsultations increase guideline adherence and treatment quality in infectious disease and intensive care management, creating added value for critically ill patients.

Trial Registration: ClinicalTrials.gov NCT03137589; <https://clinicaltrials.gov/ct2/show/NCT03137589>

(*J Med Internet Res* 2022;24(3):e34098) doi:[10.2196/34098](https://doi.org/10.2196/34098)

KEYWORDS

telemedicine; infectious disease medicine; sepsis; evidence-based medicine; eHealth

Introduction

Background

Worldwide, health workforce shortages are a pressing concern. By 2030, it is estimated that there will be a shortage of 9.9 million physicians and nurses worldwide [1,2]. In addition, the proportion of older people in Europe will exceed 30% by 2050 [3,4]. Hence, health care systems must become more flexible and efficient, for example, through accelerated digitization. Infectious disease management, especially the management of sepsis, is one area in which the potential for digitization to reduce the global burden of disease is particularly strong.

Increasing antimicrobial resistance (AMR) poses a growing threat to patients. Important underlying drivers are the overuse and misuse of antibiotics. Every year, approximately 700,000 patients worldwide die from infections that are treatable with antibiotics [5]. Global pandemics such as COVID-19 are likely to promote the overuse of antimicrobials, thereby facilitating the further development of AMR. AMR could cost US \$100 trillion between now and 2050, with the annual mortality rate reaching 10 million over this period [5].

In 2017, there were approximately 48.9 million cases of sepsis worldwide and 11 million sepsis-related deaths (19.7% of all deaths globally) [6]. Sepsis is the most common cause of morbidity and mortality in intensive care units (ICUs) around the world. Although the sepsis-related mortality rate is continuously decreasing, it is still remarkably high (30%-50%) [7-9]. The associated costs are US \$24 billion annually in the United States alone [6].

Numerous studies have shown that adherence to clinical practice guidelines for antibiotic therapy and sepsis management is associated with improved patient outcomes [10-15]. Alarming, compliance with these guidelines is low [10,16-18]. International educational health care campaigns, such as the *Choosing Wisely* initiative, have responded to this global challenge by increasing professional awareness of evidence-based medicine [19]. In general, the *Choosing Wisely* recommendations promote

essential practices and the avoidance of unnecessary diagnostic, preventive, and therapeutic procedures [15,19].

Telemedicine has the potential to support these efforts. It facilitates direct, round-the-clock interactions between physicians who care for patients in hospitals with limited subspecialist staff and intensivists or infectious disease specialists located far away. Observational studies have demonstrated that expert teleconsultations can reduce sepsis-related mortality by approximately 25%, with a simultaneous increase in guideline adherence [12,20]. Despite decades of intensive care research worldwide, no drug or other therapeutic measure has achieved a comparable reduction.

Objective

The aim of TELnet@NRW is to establish and evaluate a telemedical inpatient-outpatient network (24/7/365) to improve the application of evidence-based medicine in infectious disease management, especially the management of sepsis. We hypothesized that the establishment of a digital network based on expert teleconsultations increases treatment quality in inpatient and outpatient care for these 2 subspecialties.

Methods

Trial Design and Ethics Approval

TELnet@NRW was a multicenter, stepped-wedge cluster randomized trial conducted at 2 university hospitals (Aachen and Muenster), 17 hospitals and 103 outpatient physicians' offices associated with 2 physician networks. The protocol is publicly available [21], and the trial was prospectively registered at ClinicalTrials.gov (NCT03137589). The Ethics Committee of the Medical Faculty of the RWTH Aachen approved the study (EK 068/17). The study was funded by the Innovation Fund of the Federal Joint Committee (February 2017 to January 2020, funding code 01 NVF16010). Independent researchers from the Department of Health Economics and Health Care Management at Bielefeld University conceptualized and performed the analyses. The reporting of this trial is in line with the CONSORT-EHEALTH (Consolidated Standards of Reporting

Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist [22].

Inpatient and Outpatient Participants

Inpatients, aged ≥ 18 years, who had *Staphylococcus aureus* bacteremia or required intensive care treatment and who provided written informed consent were eligible for inclusion.

Outpatients, aged ≥ 18 years, with a possible infectious presentation based on the International Classification of Primary Care, who provided written informed consent were eligible for enrollment. Measles vaccination rates in children were also evaluated.

Because of the complexity and diversity of possible diagnoses and the extremely high total number of patients during the entire study period, triage for study enrollment was carried out by the attending physician on-site.

We excluded minors and patients who did not formally consent to participate in the study. In addition, patients who were in a dependent or employment relationship with the sponsor or one of the investigators or the principal investigator and patients who lived in an institution as mandated by a legal or administrative order were excluded from the study.

Technical Requirements

Initially, the standardized technical requirements at each participating site were established in line with the relevant guideline recommendations [23]. This also included setting up a wireless local area network in hospitals and medical practices. A secure, privacy-compliant infrastructure was used for communication, including 2 high-encryption audio-video conference systems and the certified data exchange platforms *FallAkte Plus* (Healthcare IT-Solutions) and *ELVI* (CompuGroup Medical), which complied with the General Data Protection Regulation (EU Regulation 2016/679). Overall, our data protection measures were externally reviewed by independent data protection experts and continuously monitored throughout the study. The video conferencing infrastructure met high requirements in terms of quality, data security and portability. This infrastructure operated in a high-security, closed network (hardware virtual private network) or on dedicated lines.

Interventions: Expert Teleconsultations Plus Advanced Training Courses

The innovative telemedical network TELnet@NRW involved outpatient-inpatient cooperation. Separate facilities (hospitals of different levels of care and outpatient physicians' offices) established a new digital health care structure for North Rhine-Westphalia, Germany. TELnet@NRW provided expert knowledge from 2 university hospitals to participating hospitals and outpatient physicians through expert teleconsultations (24/7/365 availability). Expert teleconsultations were provided on request after the initial triage was carried out by the attending physician on-site. Consultants for intensive care medicine participated in key care processes 24/7, whereas infectious disease specialists were available once weekly and on demand, including participation in rounds, additional expert teleconsultations, emergency consultations, and audits of clinical

patient data. Before implementation of the expert teleconsultations, participants received advanced training courses on guideline-compliant treatment.

Study Schedule and Data Collection

All clusters went through three different study phases: During the preintervention phase, pseudonymized patient data from routine care was documented (details are provided in [Multimedia Appendices 1 and 2](#), with the preintervention phase shown in red). During the subsequent transition phase (shown in white), the required hardware infrastructure was set up at the different study sites. Then, participants received on-site training according to their cluster schedule. To familiarize participants with the new processes, expert teleconsultations were already provided during the transition phase. Data on the effects of the intervention were collected during the following intervention phase (shown in blue). Primary data were generated using standardized case reporting forms. For the analyses of influenza and measles vaccination rates, we used routine outpatient claims data from the Association of Statutory Health Insurance Physicians (AHIP).

Outcomes

The primary outcome measure was adherence to the 10 *Choosing Wisely* recommendations for infectious diseases provided by the German Society for Infectious Diseases, which are applicable to both inpatient and outpatient care; these contain 5 *Dos* and *Don'ts* for infectious disease management (for each definition, please refer to [Multimedia Appendix 3](#)) [15]. Notably, the first 2 *Dos* address important quality indicators in intensive care medicine as they are associated with lower mortality [12,20,24,25]. For improved overview, [Textbox 1](#) shows the 10 *Choosing Wisely* recommendations sorted by their applicability to the inpatient and outpatient sector.

Secondary outcome measures for the inpatient sector were rate of sepsis therapy in compliance with guidelines (in compliance with the Surviving Sepsis Campaign guidelines for the management of severe sepsis and septic shock, defined as adherence to the 3- and 6-hour sepsis bundles [9]); rate of acute respiratory distress syndrome (ARDS) therapy in compliance with guidelines (measured against the evident ventilation targets, ventilation with low ventilation volumes, and low peak pressures; with controlled ventilation; breath volume of 6 mL/kg calculated ideal body weight; positive end-expiratory pressure setting in proportion with the necessary FiO_2 ; and plateau pressure < 30 cm H_2O [26]); and ICU and sepsis-related mortality, hospital mortality, ICU length of stay (LOS) and hospital LOS, rate of patients with dialysis at discharge from the ICU, and rate of transfer transport (defined as rate of patients discharged to another hospital).

Secondary outcome measures for the inpatient sector were health-related quality of life (measured using the 36-Item Short Form Survey version 2.0 questionnaire).

In our protocol, the 3 process variables *rate of sepsis diagnosis*, *rate of ARDS diagnosis*, and *rate of undiagnosed sepsis* were listed as part of the secondary outcomes. We corrected this in the report. All process variables are now reported as such.

Textbox 1. Primary outcome sorted by sector.

Primary outcome inpatient sector: adherence to the following *Choosing Wisely* recommendations

- “*Staphylococcus aureus* bloodstream infection imperatively needs efficacious antimicrobial treatment and identification and elimination of the source of infection” [P1]
- “In critically ill patients with signs of infection, early appropriate antibiotic therapy is crucial after obtaining cultures, and treatment should be regularly re-evaluated” [P2]
- “Prescribe oral forms of highly bioavailable antimicrobial agents to patients who can reliably receive and absorb medications via the enteral route” [P5]
- “Do not treat asymptomatic bacteriuria with antibiotics” [N2]
- “Do not treat *Candida* recovered from respiratory or gastrointestinal tract specimens” [N3]
- “Do not extend the administration of prophylactic antibiotics after surgery (after the patient has left the operating room)” [N4]
- “Do not treat elevated C - reactive protein or procalcitonin levels in serum with antibiotics in patients not presenting signs or symptoms of infection” [N5]

Primary outcome outpatient sector: adherence to the following *Choosing Wisely* recommendations

- “Annual influenza vaccination should be given to individuals aged >60 years, patients with specific comorbidities, and people (eg, health care workers) who may infect vulnerable persons” [P3]
- “All children should receive the measles vaccine, and adults born after 1970 without prior documented vaccination against measles should get at least one dose of the vaccine” [P4]
- “Avoid prescribing antibiotics for uncomplicated upper respiratory tract infections including bronchitis” [N1]
- “Do not treat asymptomatic bacteriuria with antibiotics” [N2]

Randomization and Masking

Participating hospitals and outpatient physician offices were randomly assigned to 4 clusters of 4 to 5 hospitals and 4 clusters of 23 to 28 outpatient physicians’ offices ([Multimedia Appendices 1 and 2](#)). Participants were randomly assigned to one of the clusters with different start dates for the intervention phase by an independent statistician using a computer-generated random allocation sequence. It was not possible to mask the health care staff or patients, as they were involved in the delivery of the intervention.

Statistical Methods

Primary outcomes were evaluated using binary logistic regression models (primary data) and zero-or-one inflated beta regression models (AHIP data).

In the analyses of the inpatient data, we controlled for the treating hospital, patient age, and the Sequential Organ Failure Assessment (SOFA) score at enrollment in the study [27]. SOFA scores were calculated based on the data in routine clinical patient records. Missing baseline values for the different SOFA subscores were imputed, if available, by measurements within the first 3 days after enrollment (next observation carried backward) [28]. Only patients with complete SOFA scores (ie, measurements for all 6 subscores) at baseline were included in the analyses.

To differentiate the training effect and the external effect of expert teleconsultations from the direct counseling effect, the intervention group was separated into patients with and without expert teleconsultations, as not all patients who received interventions were treated with expert teleconsultations. Models for ICU and sepsis-related mortality and sepsis bundle

compliance were specified in the same manner. Effects on the LOS were estimated using linear, gamma, and log-linear regression models with the same control variables.

In the evaluation of the primary outpatient data, we controlled for the treating outpatient physician and patient age. Three different models were estimated for each outcome:

- Model 1 contained only the group variable (group) and the control variables.
- In addition, model 2 contained a count variable (n) that recorded the number of expert teleconsultations the outpatient physician had already used before the visit with the respective patient.
- Finally, model 3 contained a quadratic term of the count variable (n²) to map possible learning curves among outpatient physicians in the sense of a decreasing marginal utility of the expert teleconsultations.

In the AHIP data, influenza and measles vaccination rates were documented quarterly at the practitioner level. To isolate the effect of the intervention on the vaccination rates, we controlled for the treating practitioner, the number of patients during the quarter and seasonal or quarterly effects.

Baseline group differences were tested using odds ratios (ORs) for binary variables or 2-tailed *t* tests for metric variables. Data cleaning and statistical analyses were performed in *R* (R Foundation for Statistical Computing; version 3.6.3) using the functions *glm* for logistic models and *gamlss* for beta regression models. CIs for regression estimates were computed using the *confint* command from the *stats* package, which calculated the CIs based on profile likelihood estimation. All analyses were based on a significance level of $\alpha=.05$. The model structures are detailed in the [Multimedia Appendix 4](#).

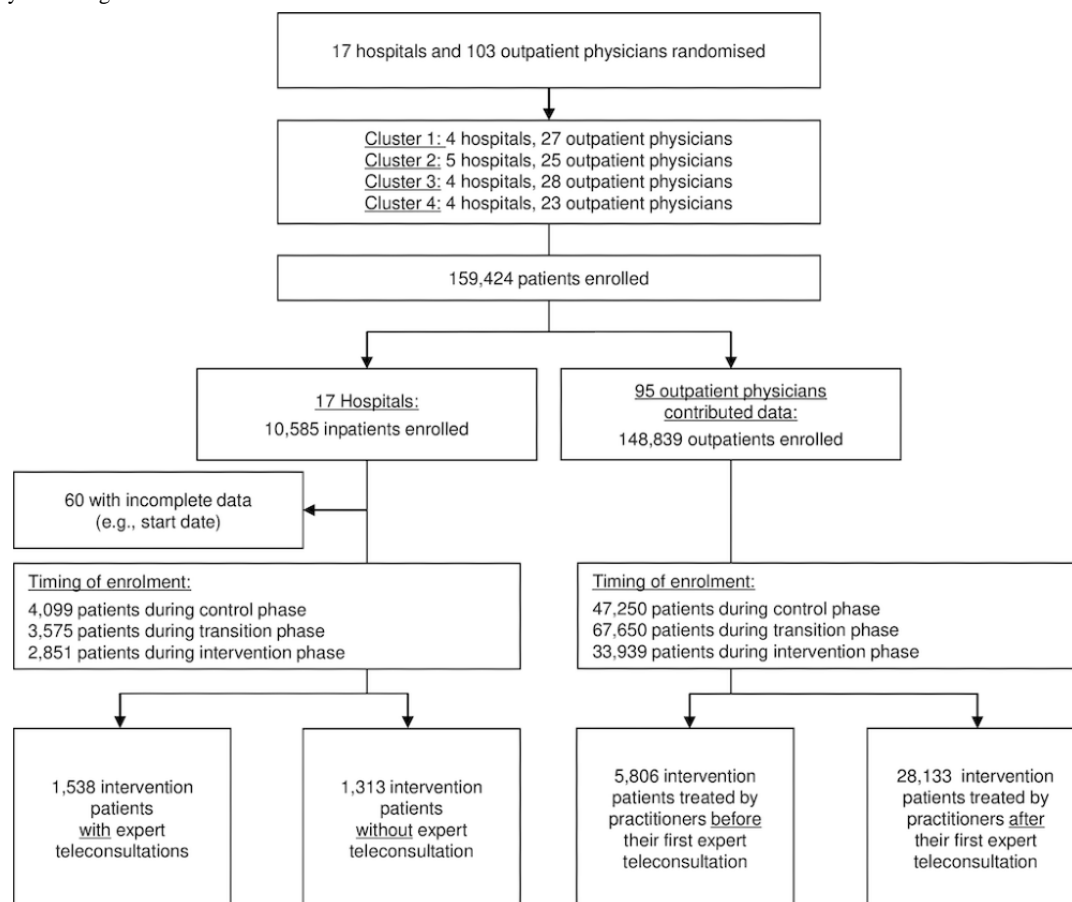
Results

Overview

A total of 17 hospitals and 103 outpatient physicians underwent randomization. The participating hospitals had between 101 and 449 beds; participating ICUs had on average 10 ICU beds (range 5-14) and were mixed medical-surgical ICUs staffed with anesthesiologists and internists. Hospitals were in both urban and rural areas. Of the 17 hospitals, 8 (47%) served populations <50,000; 7 (41%) served populations from 50,000 to 100,000; and 2 (12%) served populations >250,000. Most

outpatient physicians were in urban areas with a high outpatient physician density per capita. Among the participating doctors, multiple specialties were represented, and the majority were general practitioners, internists, ophthalmologists, or gynecologists. Between May 3, 2017, and September 30, 2019, we enrolled patients (n=159,424) who required infectious disease or intensive care treatment in our study (Figure 1). Overall, we provided 8505 inpatient expert teleconsultations. For outpatients, the average teleconsultation rate was 1.33% (1980/148,839). In the following, we report first the results of the inpatient sector and then the results of the outpatient sector.

Figure 1. Study flow diagram.



Inpatient Study Enrollment

After initial triage by the attending physicians, we enrolled 10,585 inpatients (Multimedia Appendix 5). The baseline characteristics of the inpatients are detailed in Table 1. A total of 0.57% (60/10,585) of the patients with incomplete data were excluded from the analysis, and we provide details on the distribution of missing values before and after imputation (Multimedia Appendix 6). For the recommendation *Staphylococcus aureus* bloodstream infection imperatively needs efficacious antimicrobial treatment and identification and elimination of the source of infection, the SOFA score was not

included in the statistical model, as we mainly observed non-ICU patients for this outcome, for whom the relevant parameters to calculate the score are not routinely recorded.

Compared with those in the control group, inpatients in the intervention group were older (mean age 69.25 years in the control group vs 72.14 years in the intervention group) and had higher SOFA scores at baseline (mean SOFA 3.58 in the control group vs 4.12 in the intervention group). The higher morbidity in the intervention group also manifested itself in a higher sepsis rate (5% vs 9%) and a higher ARDS rate (13% vs 17.9%). For the outcome-specific analysis samples, we found no fundamental deviations from the characteristics of the overall sample.

Table 1. Inpatient characteristics.

Inpatient characteristics	Control group	Transition group	Intervention group	Intervention group versus control group	
				Difference (95% CI)	P value
Patients, mean (SD)	4099	3575	2851	N/A ^a	N/A
Age (years), mean (SD)	69.25	70.34	72.14	2.89 (2.182-3.591)	<.001
Sex, n (%)					
Male	1920 (46.8)	1616 (45.2)	1415 (49.6)	N/A	.97
Female	2177 (53.1)	1958 (54.8)	1430 (50.2)	N/A	.97
Other	2 (0)	1 (0)	6 (0.2)	N/A	N/A
SOFA ^b score at baseline, mean (SD)	3.58	3.72	4.12	0.54 (0.408-0.687)	<.001
Sepsis incidence, n (%)	206 (5)	286 (8)	256 (9)	N/A	<.001
ARDS ^c incidence, n (%)	531 (13)	696 (19.5)	511 (17.9)	N/A	<.001

^aN/A: not applicable.

^bSOFA: Sequential Organ Failure Assessment.

^cARDS: acute respiratory distress syndrome.

Inpatient Primary Outcomes

We found significant between-group differences in the management of *Staphylococcus aureus* bloodstream infections (*Staphylococcus aureus* bloodstream infection imperatively needs efficacious antimicrobial treatment and identification and elimination of the source of infection, P1). As expert teleconsultations were provided to all patients in this analysis, we did not divide the intervention group to estimate the effects of the intervention on this recommendation. Patients in the intervention group were significantly more likely to be treated in accordance with the recommendation (OR 4.004, 95% CI 1.828-9.202; $P=.001$; [Table 2](#)).

The direct effect of expert teleconsultations became also evident in the treatment of critically ill patients with severe sepsis and septic shock. We found significant between-group differences for the recommendation *In critically ill patients with signs of infection, early appropriate antibiotic therapy is crucial after obtaining cultures, and treatment should be regularly re-evaluated* (OR 6.822, 95% CI 1.271-56.607; $P=.04$; [Table 2](#)). However, patients in the intervention group who did not receive expert teleconsultations also received treatment that was more in line with the guideline recommendation than that received by the control group. Notably, across all included patients diagnosed with severe sepsis and septic shock, adherence to this recommendation was negatively associated with patient age ([Table 2](#)).

We found no significant intervention effects for the recommendation *Prescribe oral forms of highly bioavailable antimicrobial agents to patients who can reliably receive and absorb medications via the enteral route* ([Table 2](#)).

Regarding the recommendations *Do not treat asymptomatic bacteriuria with antibiotics* ($n=24$) and *Do not treat Candida recovered from respiratory or gastrointestinal tract specimens* ($n=32$), we had too few observations to generate logistic regression models ([Table 2](#)).

Regarding the extension of the period of treatment with prophylactic antibiotics after surgery once a patient has left the operating room, we observed a higher guideline adherence in patients in the intervention group who did not receive expert teleconsultations than in patients in the control group (OR 9.372, 95% CI 1.519-111.467; $P=.04$; [Table 2](#)). However, the estimated coefficient for the portion of the intervention group who directly received expert teleconsultations remained statistically nonsignificant.

Furthermore, no significant intervention effects could be found for the recommendation *Do not treat elevated C-reactive protein or procalcitonin levels in serum with antibiotics in patients not presenting signs or symptoms of infection* ([Table 2](#)). It should be noted that the latter was already fulfilled in 90% (531/590) of the control cases.

Table 2. Regression analyses of inpatient primary outcomes.^a

	P1 ^b (N=186)			P2 ^c (N=211)			P5 ^d (N=126)			N4 ^e (N=193)			N5 ^f (N=919)		
	Compliance, % (n/N)	OR ^g (95% CI)	P value	Compliance, % (n/N)	OR (95% CI)	P value	Compliance, % (n/N)	OR (95% CI)	P value	Compliance, % (n/N)	OR (95% CI)	P value	Compliance, % (n/N)	OR (95% CI)	P value
Control variables															
SOFA ^h score	— ⁱ	—	—	N/A ^j	0.973 (0.863-1.096)	.65	N/A	1.355 (1.064-1.787)	.02	N/A	1.164 (0.753-1.879)	.51	N/A	0.772 (0.608-0.975)	.03
Age	N/A	0.973 (0.944-1.000)	.06	N/A	0.952 (0.914-0.987)	.01	N/A	0.995 (0.953-1.042)	.82	N/A	1.048 (1.008-1.093)	.02	N/A	0.993 (0.979-1.007)	.34
Group variables															
Control group	16.3 (15/92)	Ref ^k	N/A	49.2 (29/59)	Ref	N/A	21.6 (19/88)	Ref	N/A	85.3 (110/129)	Ref	N/A	90.0 (531/590)	Ref	N/A
Intervention group	45.7 (43/94)	4.004 (1.828-9.202)	<.001	—	—	—	—	—	—	—	—	—	—	—	—
Without teleconsultation	—	—	—	81.8 (108/132)	4.718 (2.032-11.563)	<.001	0.00 (0/14)	0.000 (0.000, 1.032)	.99	93.9 (31/33)	9.372 (1.519-111.467)	.04	84.9 (163/192)	0.990 (0.542-1.834)	.97
With teleconsultation	—	—	—	90.0 (18/20)	6.822 (1.271-56.607)	.04	25.0 (6/24)	1.135 (0.179, 7.493)	.89	80.6 (25/31)	1.744 (0.326, 12.861)	.54	92.1 (125/137)	1.463 (0.666-3.416)	.36

^aEach model also controlled for hospital specific effects, which are not reported individually in this table; CIs were calculated based on profile likelihood estimation.

^bPrimary outcome P1: Imperatively start antimicrobial treatment and remove the focus on *Staphylococcus aureus* bloodstream infection.

^cPrimary outcome P2: Critically ill patients with signs of infection need early appropriate antibiotic therapy.

^dPrimary outcome P5: Prefer oral formulations of highly bioavailable antimicrobials whenever possible.

^ePrimary outcome N4: Do not prolong prophylactic administration of antibiotics in patients after they have left the operating room.

^fPrimary outcome N5: Do not treat an elevated C - reactive protein or procalcitonin level with antibiotics in patients without signs of infection.

^gOR: odds ratio.

^hSOFA: Sequential Organ Failure Assessment.

ⁱDue to differences model specifications, the respective variables were not included in all models.

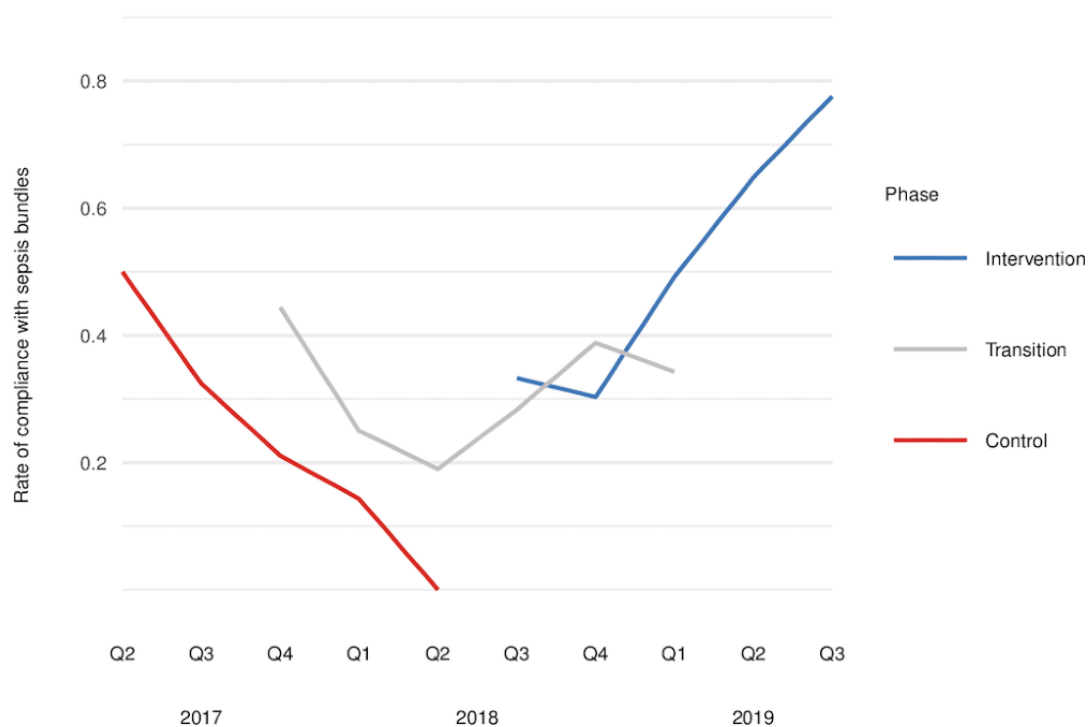
^jN/A: not applicable.

^kRef: reference group.

Inpatient Secondary Outcomes

Significant between-group differences were also found in the adherence to the 3- and 6-hour sepsis bundles for patients with the diagnoses of severe sepsis and septic shock (Figure 2 and Multimedia Appendix 7). Overall (0-6 hours), patients in the intervention group with expert teleconsultations outperformed the controls with regard to sepsis bundle compliance (OR 7.739, 95% CI 2.379-28.026; $P=.001$). This was mainly driven by an

improvement in compliance with the 4 to 6-hour bundle. With expert teleconsultations, the odds of being treated in accordance with the guideline recommendations in the 4- to 6-hour therapy course after receiving a diagnosis of sepsis were 14.2 times higher (OR 14.245, 95% CI 3.121-85.424; $P=.001$) than in the control group. Overall and for the 3- and 6-hour sepsis bundles, we found significant improvements in compliance for patients in the intervention group who did not receive direct telemedical treatment support.

Figure 2. Sepsis bundle compliance over time.

Regarding treatment quality in patients with mild ARDS, the descriptive analysis showed an increase in compliance from 7.4% (16/217) in the control group to 18.4% (9/49) in the intervention group without expert teleconsultations and 11.8% (19/161) in the group with expert teleconsultations. In the logistic regression model, a significant OR of 3.621 (95% CI 1.256-10.319; $P=.02$) was obtained for patients without expert teleconsultations. Patients with expert teleconsultations also showed a significantly increased chance of correct treatment mild ARDS (OR 2.355, 95% CI 1.023-5.516; $P=.04$). For patients with moderate or severe ARDS, we could not demonstrate these effects. A detailed description of this outcome is presented in [Multimedia Appendix 8](#).

Regarding ICU mortality (OR 1.276, 95% CI 0.909-1.794; $P=.16$) and sepsis-related mortality (OR 0.680, 95% CI 0.230-10.874; $P=.37$), no statistically significant intervention effects were found in our model estimations; however, a reduction in the sepsis-related mortality rate of 5% was achieved (19/66, 28.8% in the control group vs 50/210, 23.8% in the intervention group). Also, no significant improvements with regard to hospital mortality were observed.

ICU LOS was significantly longer for intervention patients with (+1.971 days, 95% CI 1.858-27.708; $P=.004$) and without (+2.253 days, 95% CI 2.235-40.535; $P=.002$) expert teleconsultations than for the respective controls. Regarding hospital LOS, it is noticeable that patients in the control group were hospitalized longer (mean 16.3 days, 95% CI 15.65-16.97 days) than patients in the intervention group without expert teleconsultations (mean 14.15 days, 95% CI 12.96-15.35 days) but for shorter periods than patients with expert teleconsultations (mean 20.62 days, 95% CI 19.55-21.70 days). The linear regression model confirms this result for patients with expert

teleconsultations. These patients stayed on average 4.6 days ($\beta=4.610$ days, 95% CI 3.316-5.905 days; $P<.001$) longer in hospital than control patients. For the group of patients without expert teleconsultations, there was no significant difference to the control group.

A total of 0.3% (6/1983) of the patients in the control group and 1.1% (8/739) in the intervention group were discharged from hospital on dialysis. With this very low prevalence, effects of any intervention cannot be shown.

Overall, 4.38% (86/1965) of the patients in the control group were transferred to another hospital during our study. In the intervention group, this proportion was 11.34% (143/1261). Patients with expert teleconsultations were transferred more frequently (101/857, 11.8%) than patients in the intervention group without expert teleconsultations (42/404, 10.4%). The model calculation also shows a significant intervention effect. Patients who received an expert teleconsultation had a 2.9-fold higher chance of being transferred (OR 2.903, 95% CI 2.012-4.186; $P<.001$). In patients in the intervention group without expert teleconsultations, this chance was also significantly increased compared with that in the control group (OR 2.432, 95% CI 1.570-3.721). Overall, the analyses thus show an increase in the number of transfers owing to the intervention.

Outpatient Study Enrollment

In the outpatient sector, 148,839 patients were enrolled in our study. The intervention group differed significantly from the control group with regard to their distribution between the 2 physician networks. Baseline characteristics are detailed in [Table 3](#), and [Multimedia Appendix 9](#) provides details concerning study enrollment.

Table 3. Outpatient characteristics.

Outpatient characteristics	Control group	Transition group	Intervention group	Intervention group versus control group	
				Difference (95% CI)	P value
Patients, mean (SD)	47,250	67,650	33,939	N/A ^a	N/A
Age (years), mean (SD)	42.08	40.50	42.20	0.12 (−0.4538 to 0.2208)	.49
Sex, n (%)					
Male	25,908 (54.8)	36,962 (54.6)	18,584 (54.8)	N/A	.83
Female	21,342 (45.2)	30,688 (45.4)	15,355 (45.2)	N/A	.83
Physician network, n (%)					
MuM ^b	31,248 (66.1)	43,412 (64.2)	21,388 (63)	N/A	<.001
GKS ^c	16,002 (33.9)	24,238 (35.8)	12,551 (37)	N/A	<.001

^aN/A: not applicable.

^bMuM: Medizin und Mehr eG.

^cGKS: Gesundheitsnetz Köln-Süd eV.

Outpatient Primary Outcomes

Use of expert teleconsultation was associated with a significantly higher degree of adherence to the guideline recommendations for antibiotic therapy. Overall, the treatment of patients in the intervention group was significantly more compliant than that in the control group with regard to the *Choosing Wisely* recommendation for the treatment of uncomplicated upper respiratory tract infections (N1; OR 1.343, 95% CI 1.155-1.562; $P=.001$; model 1). This effect was significantly influenced by the number of expert teleconsultations conducted, as indicated by the estimated coefficient of the count variable (OR 1.007, 95% CI 1.001-1.013; $P=.04$; model 2). In addition, the estimated coefficient for the quadratic term of the count variable showed a statistically significant negative effect, which illustrates the decreasing marginal utility of the expert teleconsultations (OR 0.9998, 95% CI 0.9996-0.9999; $P=.001$; model 3). Here, we also observed an accumulation of noncompliance in the treatment of older patients (Table 4).

Our telemedical inpatient-outpatient network also achieved significant results with regard to the management of asymptomatic bacteriuria (N2). Patients in the intervention group were more likely to be treated in line with the guideline

recommendations than were the controls (OR 9.312, 95% CI 3.794-25.936; $P<.001$; model 1). This effect was influenced more by the number of expert teleconsultations conducted than by the training of the treating physicians (OR 1.533, 95% CI 1.212-2.190; $P=.004$; model 2). Although not statistically significant, we also observed a trend toward a decreasing marginal utility of the expert teleconsultations with regard to this outcome (model 3). The *Choosing Wisely* recommendations also provide advice for increasing influenza (P3) and measles vaccinations (P4). For these analyses, we examined the quarterly vaccination rates at the physician level during the study period. In our basic model, we found no significant effect of expert teleconsultations on influenza vaccinations (rate ratio 1.089, 95% CI 0.911-1.302; $P=.34$). However, to better capture seasonal effects on the vaccination rates, we additionally constructed a model in which we extended the observation period and considered the transition phase as part of the intervention phase because the expert training occurred at the end of the control phase. In this model, intervention physicians had significantly higher influenza vaccination rates (rate ratio 1.204, 95% CI 1.079-1.344; $P=.001$). Regarding measles vaccination rates, we found no significant intervention effects in either model.

Table 4. Regression analyses of outpatient primary outcomes.^a

	Compliance, % (n/N)	Model 1		Model 2		Model 3	
		OR ^b (95% CI)	P value	OR (95% CI)	P value	OR (95% CI)	P value
N1^c (N=15,714)							
Age	N/A ^d	0.978 (0.975-0.980)	<.001	0.978 (0.975-0.980)	<.001	0.978 (0.975-0.980)	<.001
Control group	80.4 (7606/9456)	Ref ^e	N/A	Ref	N/A	Ref	N/A
Intervention group	90.2 (5643/6258)	1.343 (1.155-1.562)	<.001	1.198 (0.997-1.438)	.05	0.999 (0.806-1.238)	.99
Number of teleconsultations	— ^f	—	—	1.007 (1.001-1.013)	.03	1.032 (1.015-1.049)	<.001
Squared number of teleconsultations	—	—	—	—	—	0.9998 (0.9996-0.9999)	.001
N2^g (N=752)							
Age	N/A	0.996 (0.983-1.010)	.55	0.999 (0.985-1.012)	.83	0.999 (0.985-1.012)	.84
Control group	54.5 (145/266)	Ref	N/A	Ref	N/A	Ref	N/A
Intervention group	75.9 (369/486)	9.312 (3.794-25.936)	<.001	0.147 (0.010-1.218)	.11	0.092 (0.002-2.639)	.16
Number of teleconsultations	—	—	—	1.533 (1.212-2.190)	.004	1.717 (0.819-3.174)	.08
Squared number of teleconsultations	—	—	—	—	—	0.994 (0.978-1.038)	.65

^aEach model also controlled for physician-specific effects, which are not reported individually in this table; CIs were calculated based on profile likelihood estimation.

^bOR: odds ratio.

^cPrimary outcome N1: Avoid prescribing antibiotics for uncomplicated upper respiratory tract infections.

^dN/A: not applicable.

^eRef reference group.

^fDue to the different model specification, the respective variables were not included in all models.

^gPrimary outcome N2: Do not treat asymptomatic bacteriuria with antibiotics.

Outpatient Secondary Outcome

We had planned to assess health-related quality of life in outpatients measured with the 36-Item Short Form Survey version 2.0 questionnaire. Completed questionnaires from a total of 540 patients were available for the initial survey time t0 (study enrollment). This corresponds to 0.4% (540/148,839) of the patients included in the outpatient study. Of the 540 patients, 72 (13.3%) were eligible for a further survey after 3 or 12 months, as the contact data required for a renewed contact were collected for them at t0. The response rates in the follow-ups for these 72 patients were 32% (n=23) for t1 and 28% (n=20) for t2. An analysis of changes over time did not appear to be appropriate based on the low response rates for t1 and t2, and no analysis was performed.

Discussion

Summary of Main Findings

To the best of our knowledge, TELnet@NRW is the largest telemedical cluster randomized controlled study in Europe, with

more than 150,000 patients. We established a telemedical inpatient-outpatient network as a novel digital structure in the health care system, and we found that it measurably improved the quality of patient care. The consistent introduction and implementation of standardized communication using a certified electronic patient record was a feature that was essential for increasing the effectiveness of the processes involving the new digital health network. The key findings of this study suggest that expert teleconsultation is an effective tool to provide inpatient and outpatient physicians with evidence-based expertise on a large scale, thus improving guideline compliance and the quality of infectious disease and intensive care management.

Concerning the inpatient primary outcomes, our telemedical intervention had significant quality-improving effects on the management of *Staphylococcus aureus* bloodstream infections (P1), severe sepsis and septic shock (P2), and prophylactic antibiotic therapy (N4). Quality improvements for several outcomes reached not only those patients who were treated with direct expert teleconsultations but also other patients treated by

the same physicians. This finding can be interpreted as a positive effect of the initial training courses and/or an indirect effect of the expert teleconsultations. It can therefore be assumed that the treating physicians also apply the knowledge acquired in teleconsultations to patients whose treatment is carried out without telemedical support.

In the outpatient sector, our telemedical intervention significantly increased guideline compliance in the management of uncomplicated upper respiratory tract infections (N1) and asymptomatic bacteriuria (N2). The chance of being treated according to the recommendations was positively associated with the number of teleconsultations already participated in by the outpatient physician before the respective patient visit. Furthermore, we found evidence for a decreasing marginal utility of these teleconsultations, which may reflect the physicians' learning curve. To obtain a better understanding of physicians' learning behavior and the associated practical implications, studies with longer observational periods are needed. Basic analyses of influenza and measles vaccination rates uncovered no significant intervention effects. However, when considering the initial transition phase as part of the intervention phase, influenza vaccination in the intervention group increased significantly. This could be explained either by the better statistical control of seasonal effects (as more observation quarters were included) or by the fact that the effects were larger immediately after the start of the intervention. The latter would indicate that the transition period was probably too long to capture the full potential of our intervention. Nevertheless, as this analysis deviates from the original study plan, it can only be interpreted as exploratory.

With regard to our secondary outcomes analyzed for the inpatient sector, we found that the provision of expert teleconsultations led to a higher overall sepsis bundle adherence, which was mainly driven by improvements in compliance with the 4- to 6-hour bundle. However, although the direct relationship between sepsis bundle compliance as a quality-of-care indicator and mortality is well documented in the scientific literature [10-15,20,29], we did not observe significant intervention effects on ICU mortality, sepsis-related mortality, or hospital mortality.

Although its cross-sectoral applicability was an argument for selecting compliance with the *Choosing Wisely* recommendations as the primary outcome of this study, parts of these recommendations are not applicable to both inpatient and outpatient care, especially the ICU setting. Because gastrointestinal function is very often impaired in patients in the ICU and therefore the absorption of oral medications cannot be guaranteed, most medication is administered intravenously. Likewise, obtaining tracheal secretions and microbiological tests is a rarity in the outpatient sector, as the therapeutic consequence outside of serious infections is very low. As most patients in the ICU have an indwelling urinary catheter and are sedated, the criterion *asymptomatic* cannot be evaluated in the context of bacteriuria. Notably, the compliance with abstaining from treating elevated C-reactive protein or procalcitonin levels in the serum with antibiotics in patients without signs or symptoms of infection was already 90% (531/590) in the control phase. In summary, 40% (4/10) of *Choosing Wisely*

recommendations for infectious disease management are not fully applicable to ICU care.

Within the real-world setting of our trial, it was at the discretion of the practitioners for which patients an expert teleconsultation was requested. Hence, attending physicians tended to include patients thought to be more ill during the intervention phase of the study, which is reflected by the higher SOFA scores, sepsis, and ARDS incidences for inpatients in the intervention phase compared with those in the control phase. We controlled for such selection effects by including relevant variables in our regression models, but we could not rule out the presence of unknown confounders for which there were no data available. Although our regression models were adjusted for the SOFA score, this may not have fully controlled for the differences in the baseline risks of morbidity and mortality between the study groups. It is unclear to what extent this problem has been aggravated by the frequent occurrence of missing values for the different SOFA subscores and the associated need for data imputation. It should be highlighted that TELnet@NRW was not designed or powered to detect differences in sepsis-related mortality because the primary outcome focused on quality indicators for infectious disease management ([Multimedia Appendix 3](#)). Nevertheless, in the treatment of patients with sepsis, the early detection of sepsis followed by the early initiation of therapy conforming to the recommendations in the guidelines significantly improves the clinical outcomes [8,10,20,24,30]. Hence, if expert teleconsultations continue to be part of routine health care, we expect that this will also have a positive effect on mortality in ICUs, as has been reported in similar trials in the past [12,20,24]. The equivalent is well documented in the literature regarding adherence to evidence-based management of *Staphylococcus aureus* bacteremia. Guideline adherence significantly improves clinical outcomes and reduces mortality [31-35].

As Hemming et al [36] note, control for secular effects plays a crucial role in stepped-wedge trials. However, especially the need for an appropriately long transition phase to implement the intervention at the participating sites impeded an adequate mapping of time in our statistical models. To address this shortcoming, we estimated secular effects on our inpatient primary outcomes under consideration of transition phase data in a secondary analysis. Our models incorporate time in two ways: (1) the days since the beginning of the study (ie, the beginning of the control phase) and (2) the days since the beginning of the transition phase. Results of the estimations are displayed in the [Multimedia Appendix 10](#). For most of the primary outcomes, we observed no significant time effects (P5, N4, and N5). The estimation for P2 shows (1) a significant negative overall time effect and (2) a significant positive effect since the beginning of the implementation of the intervention. Thus, we conclude that our intervention was a major driver of the observed improvements and that our primary model without time variables may rather tend to underestimate the intervention effects. For outcome P1, however, the estimation shows significant time effects pointing in the opposite direction. It can be assumed that the negative time effect observed since the beginning of the transition phase is largely due to implementation difficulties in the participating normal wards.

However, we cannot rule out that the intervention effects measured in the primary evaluations were overestimated owing to a possible time trend.

Limitations

Overall, the findings of this study should be interpreted in the context of its limitations. The participating hospitals, ICUs, and outpatient physicians were not chosen at random; instead, the 17 sites and 103 outpatient physicians were self-selected based on their willingness to participate in a study to improve patient care. However, we chose a randomized stepped-wedge design to control for clinical characteristics, demographics, and setting to protect our findings against bias. Nevertheless, our results should be interpreted with the consideration of the possibility for selection bias due to the on-site triage of patients. We controlled for such selection effects by including relevant variables in our regression models, but we could not rule out the presence of unknown confounders for which there were no data available. We also acknowledge potential bias related to secular trends in care given the fact that the control and intervention clusters did not overlap in time. Furthermore, for some outcomes, the real effect size remained uncertain, which is reflected by the large CI. This holds true especially for sepsis bundle compliance. Nevertheless, there is sufficient certainty that the associated ORs were >1 ; thus, intervention effects existed.

Conclusions

Despite the mentioned limitations, TELnet@NRW robustly demonstrated that a cross-sectoral health network, as a new digital structure in the health care system, can develop into a quality network that operates under the premise that cross-sectoral and interregional cooperation can significantly improve evidence-based care. On the basis of the technical equipment, the principles of TELnet@NRW are transferable from intensive care medicine and infectious disease management to other subspecialist medical fields that mostly rely on expert knowledge. Thus, the concept of TELnet@NRW can be adapted to other patient populations, other conditions, or other areas, in which expertise rather than equipment needs to be transported over large distances. Our results must also be interpreted in light of the most recent SARS-CoV-2 pandemic; a digital inpatient-outpatient health network is well suited to meeting pressing challenges faced by health care systems, which we will have to address in the future (eg, staff shortages in health care sectors, lack of experts in the geographic area, and aging societies). However, further research is needed with regard to the long-term patient-relevant effects of our telemedical solution and its cost-effectiveness especially in less complex cases in the outpatient setting.

Acknowledgments

The authors thank all the patients who participated in the study and the clinical and research staff at all the participating sites, without whose assistance the study would never have been completed. The authors wish to thank all members of the TELnet@NRW study group, who are listed in [Multimedia Appendix 11](#). The authors also thank Kassenärztliche Vereinigung Nordrhein and Kassenärztliche Vereinigung Westfalen Lippe for their support of TELnet@NRW.

The funder of the study had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

GM, KS, SWL, RD, SD, DB, CJ, and AP contributed to the study concept and design. GM, KS, RD, CJ, JE, AG, CL, SWL, MH, FJ-K, SD, DB, HS-S, SR, JK, and AP contributed to data acquisition. WG, DG, SE, and AP conceptualized and performed the statistical analysis. GM, KS, RD, WG, DG, SE, SD, AP, and CB contributed to data interpretation. GM, SD, CB, DG, SE, and WG wrote the report, with critical revision for important intellectual content by all other coauthors. All authors read and approved the final manuscript. GM supervised the study.

The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflicts of Interest

GM is the coordinator of the S1 Guideline Telemedicine in Intensive Care Medicine and chairman of the German Society for Telemedicine (DGTelemed). In the past, he has given lectures for Philips. GM is cofounder of Clinomic GmbH. WG, SE, and DG report grants from the Federal Joint Committee (Gemeinsamer Bundesausschuss), during the conduct of the study. AP is a cofounder of Clinomic GmbH. The authors declare no other competing interests.

Multimedia Appendix 1

Study schedule using a stepped-wedge design (inpatient sector).

[[DOCX File, 94 KB - jmir_v24i3e34098_app1.docx](#)]

Multimedia Appendix 2

Study schedule using a stepped-wedge design (outpatient sector).

[[DOCX File , 97 KB - jmir_v24i3e34098_app2.docx](#)]

Multimedia Appendix 3

Analysis algorithms of the 10 *Choosing Wisely* recommendations.

[[DOCX File , 30 KB - jmir_v24i3e34098_app3.docx](#)]

Multimedia Appendix 4

Model structures.

[[DOCX File , 19 KB - jmir_v24i3e34098_app4.docx](#)]

Multimedia Appendix 5

Study enrollment over time (inpatient sector).

[[DOCX File , 98 KB - jmir_v24i3e34098_app5.docx](#)]

Multimedia Appendix 6

Missing value distribution of baseline Sequential Organ Failure Assessment scores.

[[DOCX File , 19 KB - jmir_v24i3e34098_app6.docx](#)]

Multimedia Appendix 7

Regression analyses of sepsis bundle compliance.

[[DOCX File , 23 KB - jmir_v24i3e34098_app7.docx](#)]

Multimedia Appendix 8

Regression analysis of acute respiratory distress syndrome therapy compliance.

[[DOCX File , 23 KB - jmir_v24i3e34098_app8.docx](#)]

Multimedia Appendix 9

Study enrolment over time (outpatient sector).

[[DOCX File , 158 KB - jmir_v24i3e34098_app9.docx](#)]

Multimedia Appendix 10

Regression analyses of secular effects on inpatient primary outcomes.

[[DOCX File , 19 KB - jmir_v24i3e34098_app10.docx](#)]

Multimedia Appendix 11

TELnet@NRW Study Group.

[[DOCX File , 18 KB - jmir_v24i3e34098_app11.docx](#)]

Multimedia Appendix 12

CONSORT-EHEALTH Checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 304 KB - jmir_v24i3e34098_app12.pdf](#)]

References

1. Global Strategy on Human Resources for Health: Workforce 2030. Geneva, Switzerland: World Health Organization; 2016.
2. Commission Staff Working document on an action plan for the EU health workforce. Publications Office of the European Union. URL: <https://op.europa.eu/de/publication-detail/-/publication/402bad92-c66f-4d45-9984-70e199dfd312/language-en> [accessed 2022-02-10]
3. Strategic research agenda for robotics in Europe 2014-2020. SPARC The Partnership for Robotics in Europe. 2013. URL: https://www.eu-robotics.net/cms/upload/topic_groups/SRA2020_SPARC.pdf [accessed 2022-01-10]
4. Population structure and ageing. Eurostat. 2021. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing [accessed 2022-02-10]
5. Tackling drug-resistant infections globally: final report and recommendations. Review on Antimicrobial Resistance. 2016. URL: https://amr-review.org/sites/default/files/160518_Final%20paper_with%20cover.pdf [accessed 2022-02-10]
6. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study. *Lancet* 2020;395(10219):200-211 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(19\)32989-7](https://doi.org/10.1016/S0140-6736(19)32989-7)] [Medline: [31954465](https://pubmed.ncbi.nlm.nih.gov/31954465/)]

7. Fleischmann C, Scherag A, Adhikari NK, Hartog CS, Tsaganos T, Schlattmann P, International Forum of Acute Care Trialists. Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations. *Am J Respir Crit Care Med* 2016;193(3):259-272. [doi: [10.1164/rccm.201504-0781OC](https://doi.org/10.1164/rccm.201504-0781OC)] [Medline: [26414292](https://pubmed.ncbi.nlm.nih.gov/26414292/)]
8. SepNet Critical Care Trials Group. Incidence of severe sepsis and septic shock in German intensive care units: the prospective, multicentre INSEP study. *Intensive Care Med* 2016 Dec;42(12):1980-1989. [doi: [10.1007/s00134-016-4504-3](https://doi.org/10.1007/s00134-016-4504-3)] [Medline: [27686355](https://pubmed.ncbi.nlm.nih.gov/27686355/)]
9. Dellinger RP, Levy MM, Rhodes A, Annane D, Gerlach H, Opal SM, Surviving Sepsis Campaign Guidelines Committee including the Pediatric Subgroup. Surviving sepsis campaign: international guidelines for management of severe sepsis and septic shock: 2012. *Crit Care Med* 2013;41(2):580-637. [doi: [10.1097/CCM.0b013e31827e83af](https://doi.org/10.1097/CCM.0b013e31827e83af)] [Medline: [23353941](https://pubmed.ncbi.nlm.nih.gov/23353941/)]
10. Levy MM, Rhodes A, Phillips GS, Townsend SR, Schorr CA, Beale R, et al. Surviving Sepsis Campaign: association between performance metrics and outcomes in a 7.5-year study. *Intensive Care Med* 2014;40(11):1623-1633. [doi: [10.1007/s00134-014-3496-0](https://doi.org/10.1007/s00134-014-3496-0)] [Medline: [25270221](https://pubmed.ncbi.nlm.nih.gov/25270221/)]
11. Rhodes A, Phillips G, Beale R, Cecconi M, Chiche JD, De Backer D, et al. The Surviving Sepsis Campaign bundles and outcome: results from the International Multicentre Prevalence Study on Sepsis (the IMPreSS study). *Intensive Care Med* 2015;41(9):1620-1628. [doi: [10.1007/s00134-015-3906-y](https://doi.org/10.1007/s00134-015-3906-y)] [Medline: [26109396](https://pubmed.ncbi.nlm.nih.gov/26109396/)]
12. Deisz R, Rademacher S, Gilger K, Jegen R, Sauerzapfe B, Fitzner C, et al. Additional telemedicine rounds as a successful performance-improvement strategy for sepsis management: observational multicenter study. *J Med Internet Res* 2019;21(1):e11161 [FREE Full text] [doi: [10.2196/11161](https://doi.org/10.2196/11161)] [Medline: [30664476](https://pubmed.ncbi.nlm.nih.gov/30664476/)]
13. Wathne JS, Harthug S, Kleppe LK, Blix HS, Nilsen RM, Charani E, et al. The association between adherence to national antibiotic guidelines and mortality, readmission and length of stay in hospital inpatients: results from a Norwegian multicentre, observational cohort study. *Antimicrob Resist Infect Control* 2019;8:63 [FREE Full text] [doi: [10.1186/s13756-019-0515-5](https://doi.org/10.1186/s13756-019-0515-5)] [Medline: [31011417](https://pubmed.ncbi.nlm.nih.gov/31011417/)]
14. Dylis A, Boureau AS, Coutant A, Batard E, Javaudin F, Berrut G, et al. Antibiotics prescription and guidelines adherence in elderly: impact of the comorbidities. *BMC Geriatr* 2019;19(1):291 [FREE Full text] [doi: [10.1186/s12877-019-1265-1](https://doi.org/10.1186/s12877-019-1265-1)] [Medline: [31664914](https://pubmed.ncbi.nlm.nih.gov/31664914/)]
15. Lehmann C, Berner R, Bogner JR, Cornely OA, de With K, Herold S, et al. The "Choosing Wisely" initiative in infectious diseases. *Infection* 2017;45(3):263-268. [doi: [10.1007/s15010-017-0997-0](https://doi.org/10.1007/s15010-017-0997-0)] [Medline: [28290130](https://pubmed.ncbi.nlm.nih.gov/28290130/)]
16. Leone M, Ragonnet B, Alonso S, Allaouchiche B, Constantin JM, Jaber S, AzuRéa Group. Variable compliance with clinical practice guidelines identified in a 1-day audit at 66 French adult intensive care units. *Crit Care Med* 2012;40(12):3189-3195. [doi: [10.1097/CCM.0b013e31826571f2](https://doi.org/10.1097/CCM.0b013e31826571f2)] [Medline: [23027124](https://pubmed.ncbi.nlm.nih.gov/23027124/)]
17. Brunkhorst FM, Engel C, Ragaller M, Welte T, Rossaint R, Gerlach H, German Sepsis Competence Network (SepNet). Practice and perception--a nationwide survey of therapy habits in sepsis. *Crit Care Med* 2008;36(10):2719-2725. [doi: [10.1097/CCM.0b013e318186b6f3](https://doi.org/10.1097/CCM.0b013e318186b6f3)] [Medline: [18766100](https://pubmed.ncbi.nlm.nih.gov/18766100/)]
18. Fact sheet antimicrobial resistance. World Health Organization. 2018. URL: <https://www.who.int/en/news-room/fact-sheets/detail/antimicrobial-resistance> [accessed 2022-02-10]
19. Choosing wisely: a special report on the first five years. ABIM Foundation. 2010. URL: <https://www.choosingwisely.org/wp-content/uploads/2017/10/Choosing-Wisely-at-Five.pdf> [accessed 2022-01-21]
20. Lilly CM, McLaughlin JM, Zhao H, Baker SP, Cody S, Irwin RS, UMass Memorial Critical Care Operations Group. A multicenter study of ICU telemedicine reengineering of adult critical care. *Chest* 2014;145(3):500-507. [doi: [10.1378/chest.13-1973](https://doi.org/10.1378/chest.13-1973)] [Medline: [24306581](https://pubmed.ncbi.nlm.nih.gov/24306581/)]
21. Telemedizinisches, intersektorales Netzwerk als neue digitale Gesundheitsstruktur zur messbaren Verbesserung der wohnortnahen Versorgung. Fakultät für Gesundheitswissenschaften, Universität Bielefeld. URL: https://www.telnet.nrw/wp-content/uploads/TELnet@NRW_Evaluationsdesign-Analyseplan.pdf [accessed 2022-01-21]
22. 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](https://doi.org/10.2196/jmir.1923)] [Medline: [22209829](https://pubmed.ncbi.nlm.nih.gov/22209829/)]
23. Telemedicine in intensive care medicine. Association of the Scientific Medical Societies in Germany. 2021. URL: <https://www.awmf.org/leitlinien/detail/II/001-034.html> [accessed 2022-01-21]
24. Lilly CM, Cody S, Zhao H, Landry K, Baker SP, McIlwaine J, University of Massachusetts Memorial Critical Care Operations Group. Hospital mortality, length of stay, and preventable complications among critically ill patients before and after tele-ICU reengineering of critical care processes. *JAMA* 2011;305(21):2175-2183. [doi: [10.1001/jama.2011.697](https://doi.org/10.1001/jama.2011.697)] [Medline: [21576622](https://pubmed.ncbi.nlm.nih.gov/21576622/)]
25. Kaasch AJ, Barlow G, Edgeworth JD, Fowler Jr VG, Hellmich M, Hopkins S, ISAC, INSTINCT, SABG, UKCIRG, and Colleagues. Staphylococcus aureus bloodstream infection: a pooled analysis of five prospective, observational studies. *J Infect* 2014;68(3):242-251 [FREE Full text] [doi: [10.1016/j.jinf.2013.10.015](https://doi.org/10.1016/j.jinf.2013.10.015)] [Medline: [24247070](https://pubmed.ncbi.nlm.nih.gov/24247070/)]
26. S3-Leitlinie Invasive Beatmung und Einsatz extrakorporaler Verfahren bei akuter respiratorischer Insuffizienz. Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften. 2017. URL: https://www.awmf.org/uploads/tx_szleitlinien/001-0211_S3_Invasive_Beatmung_2017-12.pdf [accessed 2022-01-21]

27. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. *Intensive Care Med* 1996;22(7):707-710. [doi: [10.1007/BF01709751](https://doi.org/10.1007/BF01709751)] [Medline: [8844239](https://pubmed.ncbi.nlm.nih.gov/8844239/)]
28. Kenward MG, Molenberghs G. Last observation carried forward: a crystal ball? *J Biopharm Stat* 2009;19(5):872-888. [doi: [10.1080/10543400903105406](https://doi.org/10.1080/10543400903105406)] [Medline: [20183449](https://pubmed.ncbi.nlm.nih.gov/20183449/)]
29. Ferrer R, Artigas A, Levy MM, Blanco J, González-Díaz G, Garnacho-Montero J, Edusepsis Study Group. Improvement in process of care and outcome after a multicenter severe sepsis educational program in Spain. *JAMA* 2008;299(19):2294-2303. [doi: [10.1001/jama.299.19.2294](https://doi.org/10.1001/jama.299.19.2294)] [Medline: [18492971](https://pubmed.ncbi.nlm.nih.gov/18492971/)]
30. Dellinger RP, Carlet JM, Masur H, Gerlach H, Calandra T, Cohen J, et al. Surviving Sepsis Campaign guidelines for management of severe sepsis and septic shock. *Intensive Care Med* 2004;30(4):536-555. [doi: [10.1007/s00134-004-2210-z](https://doi.org/10.1007/s00134-004-2210-z)] [Medline: [14997291](https://pubmed.ncbi.nlm.nih.gov/14997291/)]
31. Goto M, Schweizer ML, Vaughan-Sarrazin MS, Perencevich EN, Livorsi DJ, Diekema DJ, et al. Association of evidence-based care processes with mortality in staphylococcus aureus bacteremia at veterans health administration hospitals, 2003-2014. *JAMA Intern Med* 2017;177(10):1489-1497 [FREE Full text] [doi: [10.1001/jamainternmed.2017.3958](https://doi.org/10.1001/jamainternmed.2017.3958)] [Medline: [28873140](https://pubmed.ncbi.nlm.nih.gov/28873140/)]
32. Pérez-Rodríguez MT, Sousa A, López-Cortés LE, Martínez-Lamas L, Val L, Baroja A, et al. Moving beyond unsolicited consultation: additional impact of a structured intervention on mortality in Staphylococcus aureus bacteraemia. *J Antimicrob Chemother* 2019;74(4):1101-1107. [doi: [10.1093/jac/dky556](https://doi.org/10.1093/jac/dky556)] [Medline: [30689894](https://pubmed.ncbi.nlm.nih.gov/30689894/)]
33. López-Cortés LE, Del Toro MD, Gálvez-Acebal J, Bereciartua-Bastarrica E, Fariñas MC, Sanz-Franco M, REIPI/SAB group. Impact of an evidence-based bundle intervention in the quality-of-care management and outcome of Staphylococcus aureus bacteremia. *Clin Infect Dis* 2013;57(9):1225-1233. [doi: [10.1093/cid/cit499](https://doi.org/10.1093/cid/cit499)] [Medline: [23929889](https://pubmed.ncbi.nlm.nih.gov/23929889/)]
34. Nagao M, Yamamoto M, Matsumura Y, Yokota I, Takakura S, Teramukai S, et al. Complete adherence to evidence-based quality-of-care indicators for Staphylococcus aureus bacteremia resulted in better prognosis. *Infection* 2017;45(1):83-91. [doi: [10.1007/s15010-016-0946-3](https://doi.org/10.1007/s15010-016-0946-3)] [Medline: [27709434](https://pubmed.ncbi.nlm.nih.gov/27709434/)]
35. van Hal SJ, Jensen SO, Vaska VL, Espedido BA, Paterson DL, Gosbell IB. Predictors of mortality in Staphylococcus aureus bacteremia. *Clin Microbiol Rev* 2012;25(2):362-386 [FREE Full text] [doi: [10.1128/CMR.05022-11](https://doi.org/10.1128/CMR.05022-11)] [Medline: [22491776](https://pubmed.ncbi.nlm.nih.gov/22491776/)]
36. Hemming K, Taljaard M, McKenzie JE, Hooper R, Copas A, Thompson JA, et al. Reporting of stepped wedge cluster randomised trials: extension of the CONSORT 2010 statement with explanation and elaboration. *BMJ* 2018;363:k1614 [FREE Full text] [doi: [10.1136/bmj.k1614](https://doi.org/10.1136/bmj.k1614)] [Medline: [30413417](https://pubmed.ncbi.nlm.nih.gov/30413417/)]

Abbreviations

AHIP: Association of Statutory Health Insurance Physicians

AMR: antimicrobial resistance

ARDS: acute respiratory distress syndrome

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

ICU: intensive care unit

LOS: length of stay

OR: odds ratio

SOFA: Sequential Organ Failure Assessment

Edited by A Mavragani; submitted 06.10.21; peer-reviewed by M Ohl; comments to author 16.11.21; revised version received 10.01.22; accepted 22.01.22; published 02.03.22.

Please cite as:

Marx G, Greiner W, Juhra C, Elkenkamp S, Gensorowsky D, Lemmen S, Englbrecht J, Dohmen S, Gottschalk A, Haverkamp M, Hempen A, Flügel-Bleienheuft C, Bause D, Schulze-Steinen H, Rademacher S, Kistermann J, Hoch S, Beckmann HJ, Lanckohr C, Lowitsch V, Peine A, Juzek-Kuepper F, Benstoem C, Sperling K, Deisz R

An Innovative Telemedical Network to Improve Infectious Disease Management in Critically Ill Patients and Outpatients (TELnet@NRW): Stepped-Wedge Cluster Randomized Controlled Trial

J Med Internet Res 2022;24(3):e34098

URL: <https://www.jmir.org/2022/3/e34098>

doi: [10.2196/34098](https://doi.org/10.2196/34098)

PMID: [35103604](https://pubmed.ncbi.nlm.nih.gov/35103604/)

©Gernot Marx, Wolfgang Greiner, Christian Juhra, Svenja Elkenkamp, Daniel Gensorowsky, Sebastian Lemmen, Jan Englbrecht, Sandra Dohmen, Antje Gottschalk, Miriam Haverkamp, Annette Hempen, Christian Flügel-Bleienheuft, Daniela Bause, Henna Schulze-Steinen, Susanne Rademacher, Jennifer Kistermann, Stefan Hoch, Hans-Juergen Beckmann, Christian Lanckohr, Volker Lowitsch, Arne Peine, Fabian Juzek-Kuepper, Carina Benstoem, Kathrin Sperling, Robert Deisz. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 02.03.2022. 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.

Original Paper

Effectiveness, Cost-effectiveness, and Cost-Utility of a Digital Smoking Cessation Intervention for Cancer Survivors: Health Economic Evaluation and Outcomes of a Pragmatic Randomized Controlled Trial

Ajla Mujcic^{1,2}, MSc; Matthijs Blankers^{2,3,4}, PhD; Brigitte Boon^{5,6,7}, PhD; Irma M Verdonck-de Leeuw^{8,9}, PhD; Filip Smit^{2,9,10}, PhD; Margriet van Laar², PhD; Rutger Engels¹, PhD

¹Erasmus School of Social and Behavioural Sciences, Erasmus University, Rotterdam, Netherlands

²Trimbos Institute, Utrecht, Netherlands

³Department of Psychiatry, Amsterdam University Medical Center, Location AMC, University of Amsterdam, Amsterdam, Netherlands

⁴Department of Research, Arkin Mental Health Care, Amsterdam, Netherlands

⁵Academy het Dorp, Arnhem, Netherlands

⁶Siza, Arnhem, Netherlands

⁷Tranzo, Tilburg University, Tilburg, Netherlands

⁸Department of Otolaryngology—Head and Neck Surgery, Amsterdam University Medical Center, Cancer Center Amsterdam, Amsterdam, Netherlands

⁹Department of Clinical, Neuro and Developmental Psychology, Amsterdam Public Health Research Institute, Vrije Universiteit, Amsterdam, Netherlands

¹⁰Department of Epidemiology and Biostatistics, Amsterdam Public Health Research Institute, Amsterdam University Medical Center, Location VUmc, Amsterdam, Netherlands

Corresponding Author:

Ajla Mujcic, MSc

Erasmus School of Social and Behavioural Sciences

Erasmus University

Burgemeester Oudlaan 50

Rotterdam, 3062PA

Netherlands

Phone: 31 30 29 59 256

Email: amujcic@trimbos.nl

Abstract

Background: Smoking cessation (SC) interventions may contribute to better treatment outcomes and the general well-being of cancer survivors.

Objective: This study aims to evaluate the effectiveness, cost-effectiveness, and cost-utility of a digital interactive SC intervention compared with a noninteractive web-based information brochure for cancer survivors.

Methods: A health economic evaluation alongside a pragmatic 2-arm parallel-group randomized controlled trial was conducted with follow-ups at 3, 6, and 12 months. The study was conducted in the Netherlands over the internet from November 2016 to September 2019. The participants were Dutch adult smoking cancer survivors with the intention to quit smoking. In total, 165 participants were included and analyzed: 83 (50.3%) in the MyCourse group and 82 (49.7%) in the control group. In the intervention group, participants had access to a newly developed, digital, minimally guided SC intervention (MyCourse-Quit Smoking). Control group participants received a noninteractive web-based information brochure on SC. Both groups received unrestricted access to usual care. The primary outcome was self-reported 7-day smoking abstinence at the 6-month follow-up. Secondary outcomes were quality-adjusted life years gained, number of cigarettes smoked, nicotine dependence, and treatment satisfaction. For the health economic evaluation, intervention costs, health care costs, and costs stemming from productivity losses were assessed over a 12-month horizon.

Results: At the 6-month follow-up, the quit rates were 28% (23/83) and 26% (21/82) in the MyCourse and control groups, respectively (odds ratio 0.47, 95% CI 0.03-7.86; $P=.60$). In both groups, nicotine dependence scores were reduced at 12 months, and the number of smoked cigarettes was reduced by approximately half. The number of cigarettes decreased more over time, and the MyCourse group demonstrated a significantly greater reduction at the 12-month follow-up (incidence rate ratio 0.87;

95% CI 0.76-1.00; $P=.04$). Intervention costs were estimated at US \$193 per participant for the MyCourse group and US \$74 for the control group. The mean per-participant societal costs were US \$25,329 (SD US \$29,137) and US \$21,836 (SD US \$25,792), respectively. In the cost-utility analysis, MyCourse was not preferred over the control group from a societal perspective. With smoking behavior as the outcome, the MyCourse group led to marginally better results per reduced pack-year against higher societal costs, with a mean incremental cost-effectiveness ratio of US \$52,067 (95% CI US \$32,515-US \$81,346).

Conclusions: At 6 months, there was no evidence of a differential effect on cessation rates; in both groups, approximately a quarter of the cancer survivors quit smoking and their number of cigarettes smoked was reduced by half. At 12 months, the MyCourse intervention led to a greater reduction in the number of smoked cigarettes, albeit at higher costs than for the control group. No evidence was found for a differential effect on quality-adjusted life years.

Trial Registration: The Netherlands Trial Register NTR6011; <https://www.trialregister.nl/trial/5434>

International Registered Report Identifier (IRRID): RR2-10.1186/s12885-018-4206-z

(*J Med Internet Res* 2022;24(3):e27588) doi:[10.2196/27588](https://doi.org/10.2196/27588)

KEYWORDS

smoking cessation; cancer survivors; effectiveness; cost-effectiveness; eHealth

Introduction

Background

In cancer survivors, continued tobacco use is one of the most important risk factors for the development of secondary cancers, iatrogenic effects of cancer treatment, and cancer mortality [1]. The prevalence of smoking among cancer survivors is considerable, estimated at 11.8% for US cancer survivors in 2018 [2], with rates that tend to be higher among women and younger cancer survivors [3,4] and those with low health-related quality of life [5]. In the Netherlands, no difference in smoking prevalence was found between cancer survivors and noncancer survivors after adjusting for sociodemographic variables [6].

Many cancer centers in the United States have not implemented tobacco treatment services [7]; less than half of cancer care providers routinely discuss smoking cessation (SC) medication with cancer survivors [8]; and the delivery of effective SC support to cancer survivors is currently lacking [9,10]. In Europe, the general picture is comparable [11]. At the same time, cancer survivors are generally receptive toward discussions of SC with their health care professionals [3,12,13]. Among patients with head and neck cancer receiving SC counseling, 26% higher SC rates were observed than control groups in a meta-analysis of 3 randomized controlled trials (RCTs) and 3 cohort studies [14]. Distance-based SC support was also found to be more effective in reducing smoking than a range of control conditions [15]. Nayan et al [16] reported that SC interventions delivered in the perioperative period lead to higher quit rates in cancer survivors (odds ratio [OR] 2.31) but found no effect of SC interventions delivered in the cancer clinic. In addition, when considering biochemically validated smoking status, no significant effect of SC interventions was found in cancer survivors [17]. An integrated tobacco treatment program in a cancer setting showed that high abstinence rates of 45.8% at 6 months could be achieved, as demonstrated in a cohort study of 3245 patients (593 had no cancer history) [18], but this was

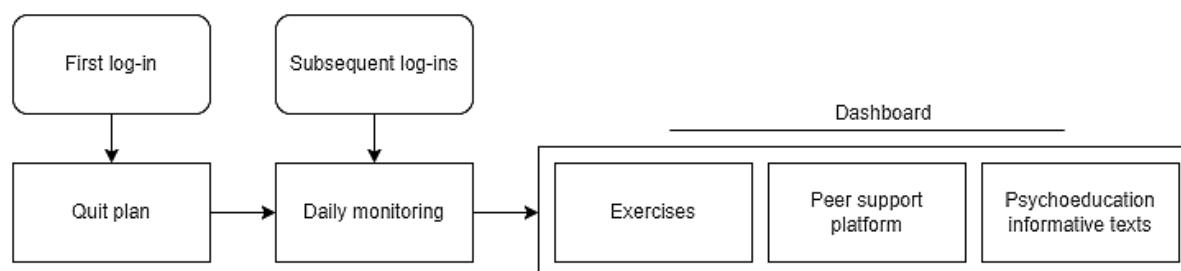
a highly intensive treatment program consisting of in person and telephone sessions spanning 8 to 12 weeks, which not only provided behavioral counseling for SC but also pharmacotherapy and treatment of related mental health conditions. Overall, there is a paucity of literature on SC interventions, specifically for cancer survivors, and the relevant literature shows conflicting outcomes.

Even fewer studies have evaluated the cost-effectiveness of digital SC interventions in the population of cancer survivors. Digital interventions may have the benefit of being scalable, easily accessible, and providing a cost-effective way to support the growing number of cancer survivors [19]. A pilot study demonstrated good acceptability of a digital SC intervention among cancer survivors [20]. A recent meta-analysis [15] indicated that few SC interventions for cancer survivors were digital interventions (2 out of 10), with most being delivered over the telephone. However, the effectiveness and cost-effectiveness of existing digital SC interventions over the internet when specifically tailored to cancer survivors is unclear.

Objectives

It was deemed timely and appropriate to launch a new study evaluating the effectiveness and cost-effectiveness of a recently developed digital intervention with minimal guidance aimed at supporting cancer survivors to quit smoking: MyCourse—Quit Smoking (in Dutch: MijnKoers—Stoppen met Roken; Figure 1 [21]). Details of how the intervention was developed are provided elsewhere [21]. In this study, we aim to answer the following research questions:

1. Is the digital interactive SC intervention *MyCourse—Quit Smoking* more effective than a web-based SC brochure to improve smoking cessation rates?
2. Is the digital interactive SC intervention *MyCourse—Quit Smoking* more cost-effective than a web-based SC brochure in terms of incremental costs per reduced pack-year and incremental costs per quality-adjusted life year (QALY) gained?

Figure 1. Intervention flowchart (adapted from Mujcic et al [21]).

Methods

Design

The effectiveness, cost-effectiveness, and cost-utility of a digital SC intervention for cancer survivors was evaluated in an individual RCT with 2 parallel arms. The trial was conducted in the Netherlands between 2016 and 2019. The first inclusion was on November 4, 2016, and the last inclusion was on September 15, 2018. The last follow-up data were collected on September 24, 2019. The study was prospectively registered in the Netherlands Trial Register (NTR6011) on September 1, 2016. For an extensive description of the study protocol, see the study by Mujcic et al [21]. This study was part of a set of 2 separate RCTs on interventions for SC and alcohol moderation, both targeting cancer survivors. The results of the RCT on the alcohol moderation intervention (MyCourse—Moderate Drinking) will be published separately.

Ethics Approval

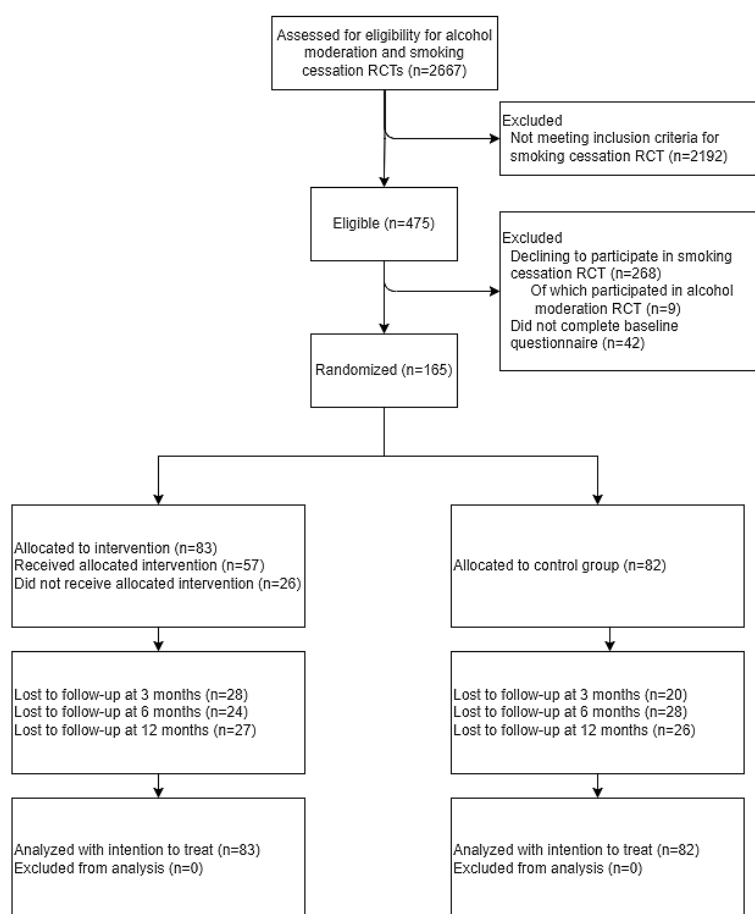
Ethical approval was obtained from an accredited medical research and ethics committee in the Netherlands (Toetsingscommissie Wetenschappelijk Onderzoek Rotterdam e.o. NL55921.101.16).

Participants and Recruitment

A dedicated website was created where participants could inform themselves about the study and apply for participation.

Applicants for the trial were eligible if they were aged ≥ 18 years, diagnosed with any form of cancer in the past 10 years, had a PC or laptop and internet connection at home, had the ability and intention to participate in the 12-month study, smoked ≥ 5 cigarettes per day in the past 7 days, and had the intention to quit smoking cigarettes. Those who had insufficient mastery of the Dutch language; those who were pregnant; or those who self-reported suicidal ideation, acute psychosis, severe alcohol dependence, dementia, or severe depression were excluded. These criteria were assessed using a web-based screening questionnaire on the website. The same screening questionnaire was used for both trials to evaluate the SC and alcohol moderation intervention [21]. Some people were eligible for both the current SC trial and the alcohol moderation trial; they were offered to participate in 1 trial of their choice (Figure 2). None of the participants were allowed to participate in both trials simultaneously.

Recruitment was conducted through web-based and offline strategies. Targeted web-based (social) media and search engine advertisements were pointed to the website and web-based screening questionnaire. SC clinics, oncology departments, and meeting centers for cancer survivors were contacted and offered promotional material (flyers and posters) to help refer cancer survivors to the website.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flowchart. RCT: randomized controlled trial.

Procedure

After completing the screening questionnaire on the study's website, applicants were informed by a computer-generated email about their eligibility for study participation. Those eligible were sent an invitation email containing patient information, an informed consent form, and a link to register. They then had 30 days to decide on their participation; during this time, they could contact the research team or an independent physician with questions. Once the informed consent form was digitally signed, the participants completed the baseline questionnaire. Immediately after completion of the baseline measurement, participants were allocated to either the intervention or the control group arm in a 1:1 ratio through adaptive randomization (minimization of baseline imbalance for age, sex, and education level) performed automatically by a server-side PHP script using a Mersenne twister random number generator. Participants received an email confirming their allocation and containing their username and instructions on how to log on. The participants were not blinded to the study group allocation. At 3, 6, and 12 months after randomization, participants received a link to the web-based questionnaire via email. Nonresponders received up to 3 reminder emails and, in case of continued nonresponse, were contacted by telephone. For each completed follow-up assessment, they were reimbursed approximately US \$30. As this was a pragmatic randomized controlled trial, in both groups, patients were allowed to use

additional support (eg, nicotine replacement therapy) if they felt they needed it.

Intervention

MyCourse—Quit Smoking is a newly developed, minimally guided, digital intervention aimed at supporting SC in cancer survivors, based on well-established therapeutic approaches: motivational interviewing, cognitive behavioral therapy, and acceptance and commitment therapy. These approaches have been incorporated into effective SC interventions in the general population [22–24]. Cancer survivors and professionals in eHealth, oncology, and SC were involved throughout the development process. The intervention was accessible through a PC, tablet, and smartphone. At first, log-in participants were guided in setting up a quit plan including a quit date, after which they gained access to 13 exercises, a web-based diary for self-monitoring of tobacco use and contextual cues, information on SC and cancer, and a peer support platform (Figure 1 [21] and Multimedia Appendix 1). Participants could choose to use the intervention whenever they wanted for the duration of the study but were encouraged to log in daily for at least 4 weeks. Elsewhere, we have provided a more extensive description of the intervention and its development [21].

Control Group

Participants in the control group received access to a noninteractive web-based information brochure on the risks of

smoking and tips on how to quit smoking, which they could access whenever they wanted by logging into the website. It contained both general SC information and information specifically relevant to cancer survivors. However, it did not contain any responsive elements of the MyCourse intervention.

Additional Support

Participants in both groups were free to seek additional help if needed and were referred to the National SC Information Line (in Dutch: *Rokeninfo-lijn*) for more information. The use of additional support was retrospectively assessed at follow-up. At the end of the study, at 12 months after randomization, the control group participants also received access to the digital intervention MyCourse—Quit Smoking.

Measures

Baseline

Sociodemographic characteristics and type of cancer were assessed. Tobacco use was assessed by using Timeline Followback (TLFB) self-reports [25] for the number of cigarettes smoked in the past 7 days. Nicotine dependence was assessed using the Fagerström Test for Nicotine Dependence (FTND) [26], a 6-item questionnaire. In participants reporting alcohol use, problematic alcohol use was assessed using the Alcohol Use Disorders Identification Test (AUDIT) [27], a 10-item questionnaire. The Marlowe-Crowne Social Desirability Scale (MCSDS) was used to assess the reliability of the self-reported questionnaire data [28]. QALYs were assessed using the 5-level EuroQol (EQ-5D-5L) [29]. In addition, the Medical Outcomes Study Short Form Survey-36 was administered to calculate the Short Form 6-dimension (SF-6D) quality of life measure [30] using the Brazier algorithm [31].

Follow-up Measurements

At all follow-up measurements, we assessed tobacco and alcohol use by using TLFB self-reports, nicotine dependence by using FTND, productivity and health care costs, QALYs using EQ-5D-5L and Medical Outcomes Study Short Form Survey-36, and the use of other SC support and e-cigarettes. Intervention use variables (eg, number of log-ins and use of major content elements) were collected throughout the study period. At 3-month follow-up treatment, satisfaction was assessed using the German adapted Client Satisfaction Questionnaire (Fragebogen zur Messung der Patientenzufriedenheit, ZUF-8), which was translated into Dutch [32].

Primary and Secondary Outcome Measures

The primary predefined endpoint was 7-day smoking abstinence at the 6-month follow-up, measured by using TLFB self-reports. Those who reported not smoking at all in the past 7 days were considered abstinent smokers (yes or o). Secondary measures included the number of smoked cigarettes in the past 7 days, nicotine dependence as measured by the FTND (range 0-10), treatment satisfaction as measured by ZUF-8 (range 8-32), health care costs, productivity loss, and QALYs.

Costs

Costs were calculated from a societal perspective for the year 2019. Intervention costs included depreciation costs (the

estimated loss of value of an interactive website, as it needs to be updated regularly to keep up with technological advancements and prevent safety issues), costs for hosting the website, and technical support and recruitment costs (which consisted of both advertising costs in web-based and offline media, as well as the costs of printing promotional material), and these were allocated evenly to all participants regardless of intervention use. Recruitment costs were included as they were considered an essential part of the intervention and control condition. Health care costs were calculated by multiplying all reported contacts with health services with the standard unit cost prices for the Netherlands [33]. Health service costs stemmed from contacts with specialized somatic and mental health care and patients' out-of-pocket costs for home care, but travel costs were not included because in both groups the interventions were delivered over the internet. Other health care costs included appointments for physiotherapy, alternative medicine, and social work. Medication costs were valued by multiplying the reported dose of that drug with unit cost price [34]. Productivity loss included costs from absenteeism and presenteeism, calculated according to the friction cost method, meaning productivity losses were limited to a maximum of 85 days after which production losses cease to exist because the sick employee would be replaced by another and using an elasticity factor of 0.8 because there is not a strict 1:1 relation between days not worked and productivity losses. Cost data related to health care use and productivity loss were assessed using the Trimbos and iMTA questionnaire for Costs associated with Psychiatric Illness [35]. Cumulative societal costs over the entire follow-up period of 12 months were calculated as the sum of health care costs and productivity losses. Costs were converted from euros to US dollars using purchasing power parities for the reference year 2019.

Sample Size

A study on an SC intervention among cancer survivors found a quit rate of 30% in the active SC intervention group versus 15% in the SC control group, translating to a relative risk (RR) of 2.14 [36]. A pilot trial of an acceptance and commitment theory web-based SC intervention found a 23% quit rate in the experimental arm versus a 10% quit rate in the control arm, translating to an RR of 2.20 [24]. On the basis of the average of these RRs, an RR of 2.1 was expected, translating into a 21% quit rate in the experimental arm, assuming a 10% quit rate in the control arm at the 6-month follow-up. On the basis of the conventional statistical significance level (Cronbach $\alpha \leq 0.05$), an RR of 2.1 at the 6-month follow-up, 204 participants would yield a power of 0.83 for 1-sided tests or a power of 0.74 for 2-sided tests.

Statistical Analyses

Imputation of Missing Data

All primary and secondary outcome measures were analyzed in accordance with the intention-to-treat principle, except for ZUF-8 (treatment satisfaction). To that end, missing data for primary and secondary outcome measures and costs were imputed using the predictive mean matching method from the mice package in R (R Foundation) [37]. The responses to ZUF-8 were not imputed. For the 2 deceased participants, the smoking

status and number of cigarettes smoked were left missing, and the quality of life (EQ-5D-5L) score and costs were set to 0.

Effect Evaluation

Tobacco abstinence (binary yes or no outcome) was analyzed using a Generalized Linear Mixed Model (GLMM) with a binomial distribution and log link function. Although imputation of missing values is not always deemed necessary when running a GLMM, imputation of missing values before running a GLMM allowed us to consider all variables that could have impacted dropout and not only the variables within a specific model. The number of cigarettes smoked (count data 0, 1, ..., N) was analyzed using a GLMM with a log link function and negative binomial distribution [38]. Included covariates were the minimized variables (gender, age, and education) and the MCSDS (social desirability of responses). Model estimates, ORs, incidence rate ratios (IRRs) or Cohen *d*, 95% CIs, and *P* values were reported. The effect of time on the number of cigarettes was analyzed using an *F* test. Differences between the intervention and control groups on FTND nicotine dependence and ZUF-8 patient satisfaction scores were analyzed using a Linear Mixed Model for the Gaussian distribution with identity as the link function; estimates, 95% CIs, and *P* values were also reported.

Cost-effectiveness Analyses

An economic evaluation was conducted alongside this RCT following the approach of Drummond et al [39] and in concordance with the Consolidated Health Economic Evaluation Reporting Standards statement [40]. QALYs over the entire follow-up period were computed using the Dutch tariff (utility weights) [41] through the area under the curve method; that is, linear interpolation for cumulating the cost over the 12-month follow-up period. The incremental cost-effectiveness ratio (ICER) was calculated as follows: $ICER = (C_1 - C_0) / (E_1 - E_0)$, where *C* refers to costs, *E* refers to effect, and the subscripts 1 and 0 refer to the MyCourse and control groups, respectively. We generated 2500 replicate samples by bootstrap and estimated the corresponding incremental costs and effects for each replicate sample, which were then plotted on a cost-effectiveness plane. In addition to the ICER per QALY gained, the ICER per reduced pack-year was calculated. Pack-years were calculated by multiplying cigarettes smoked in the past week by 52 (weeks

in a year) and dividing by 20 (cigarettes per pack) and 365 (days in a year). We calculated ICERs from the following four perspectives: societal, health care, productivity loss, and intervention cost-only. Cost-effectiveness acceptability curves (CEACs) were graphed to assess the likelihood that the intervention was deemed cost-effective, given a series of willingness-to-pay ceilings for gaining 1 QALY.

Sensitivity Analyses

The negative binomial and binomial analyses of the mice-imputed data constituted the main analyses. We conducted several sensitivity analyses for the effectiveness and incremental cost-effectiveness analyses using QALYs based on the SF-6D (instead of the EQ-5D-5L), using Winsorizing cost outliers, using the Amelia-2 package (R Foundation) instead of the mice-package for imputations, considering a gradual decline in pack-years, and using different statistical models (Multimedia Appendix 1).

Results

Sample Characteristics

The participant flow and retention rates are shown in Figure 2. Of the 2192 ineligible people, 1684 (76.82%) had no diagnosis of cancer in the past 10 years. Of the 475 eligible cancer survivors, 268 (56.4%) declined to participate, 9 (3.4%) of whom chose to participate in our parallel RCT on MyCourse for alcohol moderation, and another 42 (8.8%) cancer survivors did not complete the baseline questionnaire and were, therefore, not randomized. Sociodemographic and other characteristics are reported in Table 1. The participants' mean age was 54.2 (SD 11.2) years, 17.6% (29/165) were men, approximately half were married or living together (93/165, 56.4%), and 30.3% (50/165) had a lower education level. On average, participants had smoked for 34.5 (SD 12.0) years and smoked 100 (SD 51.2) cigarettes per week. Three participants quit smoking between screening and completing the baseline questionnaire (Table 2). Breast cancer (75/165, 45.5%), lung cancer (23/165, 13.9%), uterus cancer (19/165, 11.5%), and head and neck cancer (18/165, 10.9%) were the most frequently reported. There was no difference in the proportion of missing data between groups at any of the time points ($\chi^2_1=0.09$; *P*=.77; Multimedia Appendix 2).

Table 1. Baseline characteristics of study participants.^a

	MyCourse (n=83)	Control (n=82)	Total (N=165)
Gender, n (%)			
Women	70 (84)	66 (80)	136 (82.4)
Men	13 (16)	16 (20)	29 (17.6)
Age (years), mean (SD)	55.0 (12.1)	53.3 (10.3)	54.2 (11.2)
Education, n (%)			
Higher level	25 (30)	19 (23)	44 (26.7)
Midlevel	33 (40)	38 (46)	71 (43.0)
Lower level	25 (30)	25 (30)	50 (30.3)
Marital status, n (%)			
Married or living together	47 (57)	46 (56)	93 (56.4)
Unmarried or living alone	15 (18)	11 (13)	26 (15.8)
Divorced	16 (19)	20 (24)	36 (21.8)
Widowed	5 (6)	5 (6)	10 (6.1)
Smoking behavior, mean (SD)			
Years smoked	34.4 (11.8)	34.6 (12.2)	34.5 (12.0)
Number of cigarettes in the past 7 days	101.8 (54.3)	98.2 (48.2)	100 (51.2)
FTND ^b	4.9 (2.4)	4.9 (2.3)	4.9 (2.4)
Drinking behavior			
Drank alcohol in the last month, n (%)	55 (66)	55 (67)	110 (66.7)
Number of drinks in the past 7 days, mean (SD)	6.9 (13.1)	5.6 (8.7)	6.2 (11.2)
AUDIT, ^c mean (SD)	3.7 (5.1)	3.6 (4.2)	3.6 (4.7)
Cancer diagnosis, n (%)			
Breast	42 (51)	33 (40)	75 (45.4)
Lung	14 (17)	9 (11)	23 (13.9)
Uterus	7 (8)	12 (15)	19 (11.5)
Head and neck	10 (12)	8 (10)	18 (10.9)
Colon	5 (6)	5 (6)	10 (6.0)
Other (including bladder, lymphatic, melanoma, skin, kidney, prostate, etc)	5 (6)	26 (32)	20 (12.1)

^aPercentages may not add up to 100 because of rounding.

^bFTND: Fagerström Test for Nicotine Dependence.

^cAUDIT: Alcohol Use Disorders Identification Test.

Table 2. Smoking behavior outcomes and treatment effects (missing data were imputed; a total of 3 participants quit smoking between screening and completing the baseline questionnaire).

Variable	MyCourse (n=83)	Control (n=82)	Effect size (95% CI)
Cessation, n (%)^a			
Baseline	2 (2.4)	1 (1.2)	N/A ^b
3 months	18 (21.7)	19 (23.2)	Adjusted OR ^c 0.33 (0.02 to 5.44)
6 months	23 (27.7)	21 (25.6)	Adjusted OR 0.47 (0.03 to 7.86)
12 months	27 (32.6)	23 (28.1)	Adjusted OR 0.58 (0.03 to 9.78)
Number of cigarettes, mean (SD)^d			
Baseline	101.8 (54.3)	98.2 (48.2)	N/A ^b
3 months	54.3 (51.1)	54.2 (48.2)	Adjusted IRR ^e 0.95 (0.85 to 1.06)
6 months	50.5 (50.5)	50.1 (47.5)	Adjusted IRR 0.96 (0.85 to 1.08)
12 months	45.4 (50.9)	49.6 (44.9)	Adjusted IRR 0.87 (0.76 to 1.00) ^f
FTND,^g mean (SD)^h			
Baseline	4.9 (2.4)	4.9 (2.3)	N/A ^b
3 months	2.9 (2.5)	2.8 (2.5)	Cohen <i>d</i> =0.03 (−0.27 to 0.34)
6 months	2.6 (2.6)	2.8 (2.6)	Cohen <i>d</i> =0.07 (−0.23 to 0.38)
12 months	2.4 (2.6)	2.7 (2.5)	Cohen <i>d</i> =0.13 (−0.18 to 0.43)

^aAdjusted coefficients are based on a binomial mixed model with random intercept in which the outcome measure at follow-up is regressed on the baseline number of cigarettes, covariates, and condition.

^bN/A: not applicable.

^cOR: odds ratio.

^dAdjusted coefficients are based on a negative binomial mixed model with random intercept in which the outcome measure at follow-up is regressed on the baseline number of cigarettes, covariates, and condition.

^eIRR: incidence rate ratio.

^fAdjusted coefficients are based on a linear mixed model with random intercept in which the outcome measure at follow-up is regressed on the baseline number of cigarettes, covariates, and condition.

^gFTND: Fagerström Test for Nicotine Dependence.

^h $P < .05$ ($P = .04$).

Treatment Uptake and Satisfaction

Overall satisfaction with the SC intervention was highest in the MyCourse group (mean 21.4, SD 4.6) than in the control group (mean 17.3, SD 6.1; Cohen $d=0.77$; $t_{108.8}=4.13$; $P<.001$; [Multimedia Appendix 2](#)). Most participants in the MyCourse group logged in at least once (57/83, 69%). The number of times participants logged in was skewed, with an average of 20 (SD 61.2) and a median of 3 (range 0–384). The average time between the first and last log-in for those who logged in at least once was 105.2 (SD 157.5; median 24) days. Most reported SC support in addition to MyCourse at the 6-month follow-up was nicotine replacement therapy (control group: 25/82, 30%; MyCourse group: 14/83, 17%) and contact with a health care professional (control group: 7/82, 9%; MyCourse group: 3/83, 4%). The use of nicotine replacement therapy was reported more often (18.1% vs 30.5% at 12 months) in the control group than in the MyCourse group ($P=.02$).

Adverse Events

Two deaths occurred in the MyCourse group over the course of the study period, which was reported to the medical research and ethics committee. The cause of death was deemed to be unrelated to the study. No other adverse events were observed.

Incremental Effects

Primary Outcome

At the 6-month follow-up, 28% (23/83) of smokers had quit smoking in the MyCourse group versus 26% (21/82) in the control group ([Table 2](#)). No difference in 7-day abstinence was found between the 2 groups (adjusted OR 0.47, 95% CI 0.03–7.86; $P=.60$) when controlling for social desirability, baseline number of cigarettes used in the last week, gender, age, and education.

Secondary Outcomes

[Table 2](#) and [Figure 3](#) present the effect estimates on the secondary outcomes. In brief, the number of cigarettes smoked in the past week was significantly reduced at all follow-ups and

in both groups compared with baseline ($F_3=51.5$; $P<.001$; Table 2 and Figure 3). At 12-month follow-up, number of cigarettes was reduced by about half in both the MyCourse group, showing an average reduction of 57 cigarettes (57/101.8, 56%), and in the control group, showing an average reduction of 48 cigarettes (48/98.2, 49%; Table 2 and Figure 4). At 12-month follow-up, the reduction of number of cigarettes smoked was significantly greater in the MyCourse group than in the control group (adjusted IRR 0.87, 95% CI 0.76-1.00; $P=.04$). At 3- and 6-month follow-up, the difference in number of cigarettes between groups was not significant.

FTND scores were significantly lower at all follow-ups in both groups compared with baseline scores, whereas the time \times condition interaction was not significant (Cohen $d=0.03$; 95% CI -0.27 to 0.34 ; $P=.95$), indicating no significant difference between the groups over time.

The mean EQ-5D-5L QALYs gained in the intervention group was 0.75 (SD 0.18) and in the control group was 0.78 (SD 0.15). There was no significant effect of treatment on quality of life based on EQ-5D-5L scores ($B=-0.03$, SE 0.03; $P=.26$).

Figure 3. Percentage of quitters in both groups at baseline and during the course of the study. A total of 3 participants quit smoking between screening and completing the baseline questionnaire.

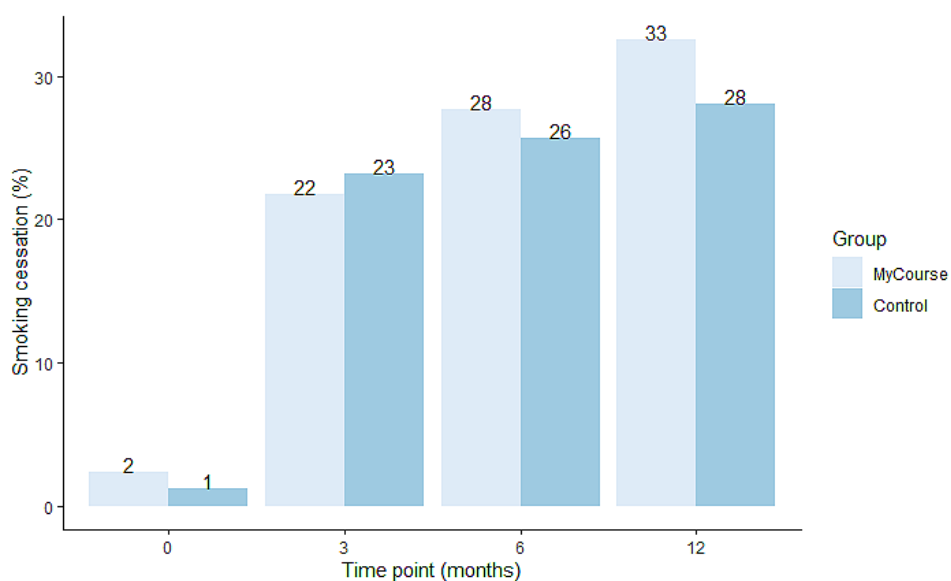
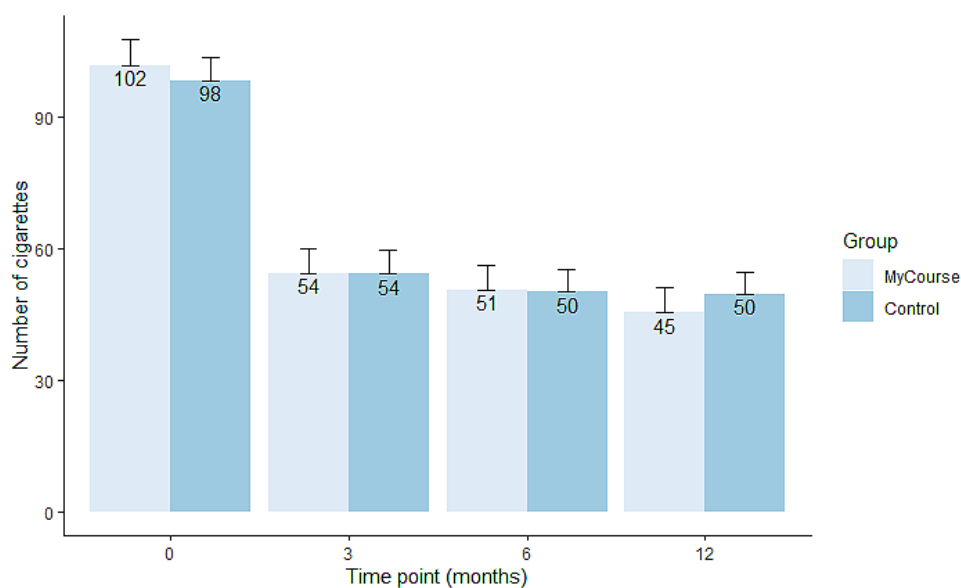


Figure 4. Mean number of cigarettes smoked in both groups at baseline and during the course of the study, including SEs. Error bars show SEs of the mean.



Incremental Costs

Table 3 presents the costs per group and the incremental costs (cost difference between the MyCourse and control groups) per cost item. The intervention costs were estimated at US \$193 per participant in the MyCourse group and US \$74 per participant in the control group. The average health care costs accumulated over the full 12 months follow-up time were US \$14,416 (SD US \$20,604) per participant in the MyCourse group and US \$12,950 (SD US \$17,704) per participant in the

control group, resulting in incremental health care costs of US \$1466 (SD US \$27,165). Cost owing to productivity losses was mainly driven by absenteeism at US \$10,444 (SD US \$17,277) in the MyCourse group and US \$8145 (SD US \$15,750) in the control group, with high within-group variance. Incremental productivity costs per participant were on average US \$1908 (SD US \$23,490). The average cumulative societal costs were US \$3493 (SD US \$38,913) higher in the MyCourse group than in the control group. See **Table 3** for a detailed breakdown of the main cost items and the corresponding SDs.

Table 3. Mean cumulative costs (in US \$) by group and incremental costs.

Cost item	MyCourse; (n=83), mean (SD)	Control; (n=82), mean (SD)	Incremental costs ^a ; mean (SD)
Health care costs	14,416 (20,604)	12,950 (17,704)	1466 (27,165)
Specialized somatic	8418 (11,792)	7180 (10,674)	1238 (15,906)
Specialized psychiatric	2151 (10,143)	1380 (5441)	771 (11,510)
Patient and family costs	1310 (10,034)	1954 (9252)	-644 (13,648)
Other	1533 (2671)	1411 (2518)	122 (3671)
Medication	1358 (3901)	1023 (3254)	335 (5080)
Productivity loss	10,720 (17,345)	8812 (15,841)	1908 (23,490)
Presenteeism	231 (733)	332 (864)	-101 (1133)
Absenteeism	10,444 (17,277)	8145 (15,750)	2299 (23,379)
Unpaid work	451 (1048)	474 (1007)	-23 (1453)
Intervention costs	193 (0)	74 (0)	119 (0)
Total societal costs	25,329 (29,137)	21,836 (25,792)	3493 (38,913)

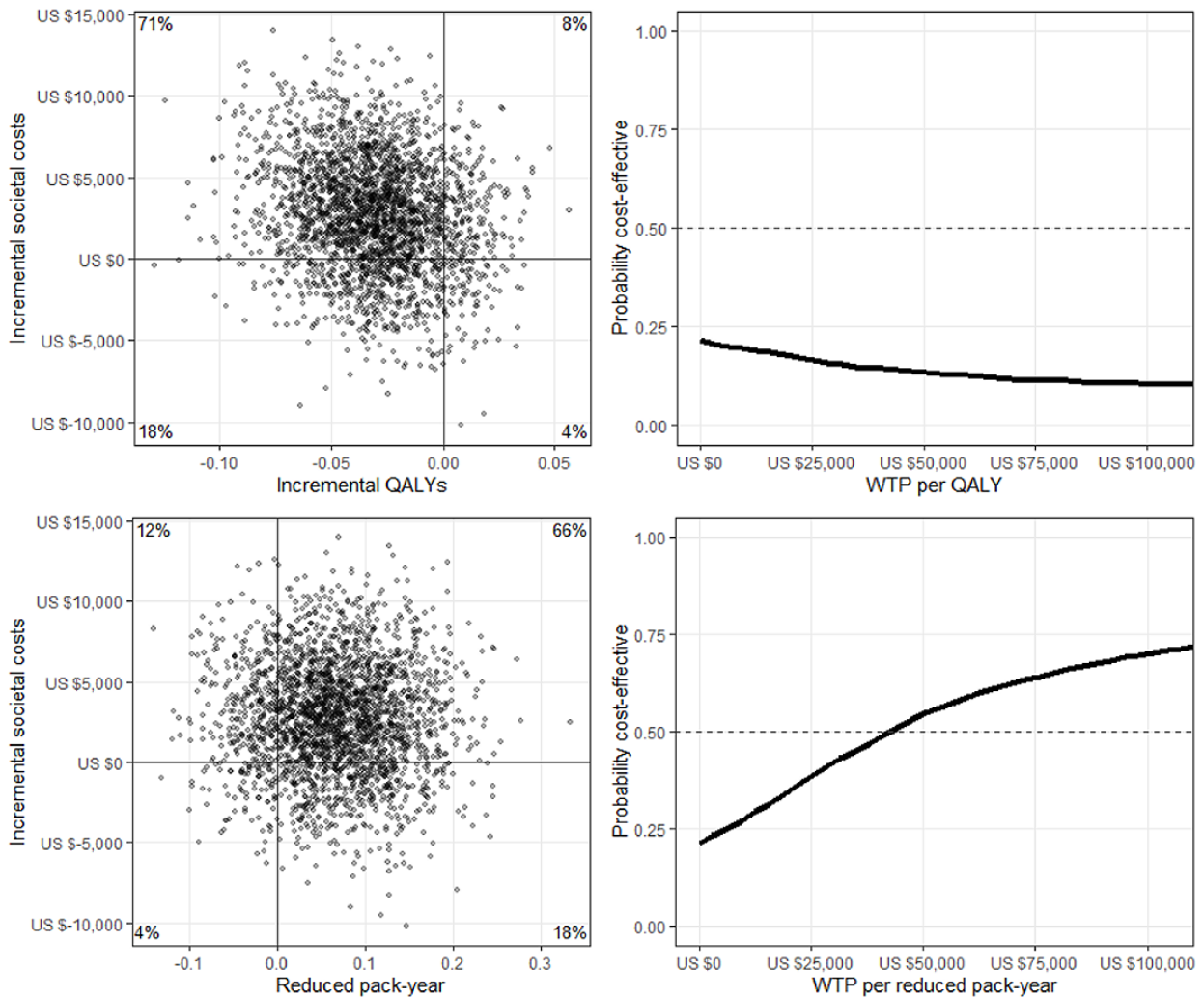
^aCosts in the MyCourse group minus costs in the control group.

Cost-Utility

Participants in the MyCourse group gained fewer QALYs than participants in the control group (0.75 vs 0.78). In addition, the societal costs in the MyCourse group were higher than those in the control group (US \$25,329 vs US \$21,836). In other words,

fewer QALYs were gained against higher costs, which rendered the control group the preferred option, as seen from a cost-effectiveness viewpoint. The cost-effectiveness plane in **Figure 5** shows that there is a 71% likelihood that MyCourse is dominated by the control group.

Figure 5. Cost-effectiveness planes and cost-effectiveness acceptability curves in US \$. Each quadrant in the cost-effectiveness planes represents a different association between the incremental costs (y-axis) and the incremental effects (x-axis) of the MyCourse group compared with the control group. When incremental cost-effectiveness ratios (ICERs) fall in the upper-right quadrant, this represents more effect at higher costs. When ICERs fall in the upper-left quadrant, this represents less effect at higher costs, meaning the MyCourse group is dominated by the control group. ICERs in the lower-right quadrant represent more effect at lower costs: the dominant quadrant. ICERs in the lower-left quadrant represent less effect at lower costs. QALY: quality-adjusted life year; WTP: willingness to pay.



Cost-effectiveness

Participants in the MyCourse group reduced pack-years more than the control group (0.41 vs 0.34) against higher societal costs (US \$25,329 vs US \$21,836). Comparing the difference in total societal costs to the difference in pack-years yielded an ICER of US \$52,067 per reduced pack-year (95% CI US \$32,515-US \$81,346). There is a 66% chance that the

intervention leads to more reduced pack-years at a higher cost (Figure 5). The CEAC based on the societal cost perspective presented in Figure 5 indicates that the intervention will be preferred over the control group when the willingness to pay per reduced pack-years is more than US \$50,000. From an intervention cost-only perspective, ICER per pack-year was calculated to be US \$1772 (95% CI US \$1384-US \$2502). See Table 4 for a breakdown by perspective.

Table 4. Incremental cost-effectiveness ratios between baseline and the 12-month follow-up.^a

Perspective	Incremental costs per reduced pack-year	
	Mean (US \$)	95% CI
Health care perspective	21,851	9179-38,920
Productivity loss perspective	28,444	16,832-44,749
Intervention cost-only perspective	1772	1384-2502
Societal perspective	52,067	32,515-81,346

^aThe incremental cost-effectiveness ratios were calculated as follows: $(C_1 - C_0) / (E_1 - E_0)$, where C refers to costs, E refers to effects, and the subscripts 0 and 1 refer to the experimental and control arms, respectively.

Sensitivity Analyses

Sensitivity analysis of the Amelia 2-imputed data (adjusted OR 3.16, 95% CI 0.17-57.80; $P=.44$) and completers only (ie, those who completed the questionnaires, without imputation; adjusted OR 2.89, 95% CI 0.16-53.18; $P=.47$) showed similar results on the effect of treatment group on tobacco abstinence rates. The Poisson model with correction for overdispersion showed similar results on the effect of treatment on cigarettes smoked at the 12-month follow-up (adjusted B=-0.055, 95% CI -1.05 to -0.04; $P=.03$). Sensitivity analyses of the Amelia-2 imputed data also showed a large effect of time but found no effect of treatment group at the 12-month follow-up (adjusted IRR 0.997, 95% CI 0.82-1.22; $P=.97$). The completers only sensitivity analysis showed a greater reduction in the number of cigarettes in the control group at 6 months (adjusted IRR 1.08, 95% CI 1.01-1.15; $P=.03$), but at 12 months, a greater reduction in the MyCourse group (adjusted IRR 0.88, 95% CI 0.82-0.94; $P<.001$). For the number of cigarettes smoked, the social desirability score was a significant predictor in the Poisson model, but not in the negative binomial model. When QALYs were based on SF-6D scores, results of the economic evaluation remained similar. When Winsorization of extreme costs was applied at the 95th percentile, the cost-effectiveness planes and CEACs remained similar, but ICER per pack-year was lower at US \$31,342 (95% CI US \$17,912-US \$50,007). When the gradual decline in the number of pack-years was accounted for, ICER per pack-year was higher (US \$68,267, 95% CI US \$42,293-US \$111,044) and the cost-effectiveness planes comparable (Multimedia Appendix 3). Overall, the sensitivity analyses attested to the robustness of the findings in the main analysis.

Discussion

Principal Findings

This study evaluated the effectiveness and cost-effectiveness of MyCourse, a digital SC intervention tailored to cancer survivors, versus a web-based noninteractive information brochure. In the MyCourse group, 27.7% of the participants quit smoking after 6 months. In the control group, 25.6% of the participants quit smoking. The number of cigarettes smoked in the past 7 days was reduced by more than half in both groups. At the 12-month follow-up, MyCourse participants showed significantly larger reductions in the number of smoked cigarettes than participants in the control group. However, no statistically significant difference was found in the SC rates

between the intervention and control groups. Nicotine dependence as measured by FTND was also significantly reduced at all time points in both groups, but no difference was found between the groups. Participants in the MyCourse group had significantly higher treatment satisfaction scores than those in the control group. From a societal perspective, the MyCourse intervention was dominated by the control group in the cost-utility analysis. In the cost-effectiveness analysis, MyCourse led to marginally better results against higher costs, with a mean ICER of US \$52,067 per reduced pack-year. Cessation rates were high in the MyCourse and control groups.

Findings in Context

We found no difference in SC rates between the MyCourse and control groups. In previous literature, digital SC interventions have shown superior effectiveness over nonactive control groups (including both usual care and printed self-help materials), among the general population as well as other target groups [42,43]. There is evidence that among cancer survivors, distance-based SC interventions (including digital interventions) also show greater effectiveness than control groups [15]. At the same time, cessation rates in the MyCourse group found in this study are comparable with cessation rates found in 2 previous studies on digital SC interventions for (childhood) cancer survivors [20,44], but cessation rates in our control group were higher than in these previous studies.

Our study did not find an effect on QALYs; a longer follow-up period may be necessary to detect improvements in quality of life among cancer survivors, as their quality of life may be more directly influenced by factors pertaining to cancer diagnosis [45]. There are also some differences in the ICERs that we found compared with previous studies. A systematic review in the Netherlands showed greater cost-utility (< US \$22,689 [€20,000] per QALY per year) of intensive SC counseling and pharmacotherapy over care as usual in another patient population: patients with chronic obstructive pulmonary disease [46], and a similar study on a digital SC intervention among the general population showed an ICER of about US \$3398 (€3000) per abstinent smoker [47]. The large ICER per reduced pack-year in this study can partly be explained by the relatively small difference in effect on the reduced number of cigarettes between the intervention and control groups, and the high health care and productivity costs in this population. Benefits or costs due to changes in tax revenue are not included in a cost-effectiveness analysis but could be a topic of interest in social cost-benefit analyses.

Cessation rates in our control group were higher than those in the studies referenced for power analysis. This might be due to differences in target groups between the studies: we studied cancer survivors, whereas the referenced studies either focused on the general population [24] or cancer survivors with a comorbid problem such as drinking or depression [36]. The similar effect on SC of the MyCourse and control groups might be due to several factors. Notably, twice as many participants in the control group reported the use of nicotine replacement therapy, suggesting that this might have influenced SC rates. The control group participants may have had an increased need for additional support. As this was a pragmatic trial, both groups were provided with the contact details of a free national telephone helpline (in Dutch: Rokeninfo-lijn), which could help find participants additional support if the current intervention was deemed insufficient. Furthermore, to recruit participants, a dedicated website and social media campaign was in place, aimed at informing cancer survivors of the short-term benefits of SC after a cancer diagnosis, emphasizing an accepting tone to reduce possible feelings of guilt, and ultimately guide them to participate in the study. Other contributing factors might have been related to the fact that over the course of the study period, participants received multiple reminder emails and telephone calls from the researchers to fill out the survey at the respective follow-up measurement waves. Although these calls were kept as short as possible, some participants might have experienced those as part of the intervention, feeling supported by them, which could have influenced SC rates. To summarize, the low-threshold provision of psychoeducation, offered in an accepting manner, encouragement to seek support provided in recruitment materials and the information brochure, repeated reminders, and increased use of nicotine replacement therapy may have been sufficient to support many participants in their SC efforts. This should be evaluated in future studies.

To considerably increase SC rates among cancer survivors compared with control groups, intensive and well-implemented programs are needed. It is possible that cancer survivors require digital interventions with more guidance, as suggested by a meta-analysis that found that only nurse-delivered SC interventions moderated effectiveness among cancer survivors [17]. Guidance might also help improve adherence, as the median number of log-ins (median 3) was lower than the recommended (almost) daily use of the intervention over the course of 4 weeks. The period between the first and last log-in came close to use as intended, with a median of 24 days.

Strengths and Limitations

The findings of this study should be interpreted in light of their strengths and limitations. A strength of this study is that the evaluation was conducted in a real-world setting: recruitment was done through both offline and web-based channels, which

will also be used for the intervention's future implementation. The difference in the number of people who completed the screening questionnaire and those who were eligible might seem to show a large selection, but this could be due to our web-based recruitment strategies, which attracted many interested people with no history of cancer on the website as well. This study recruited cancer survivors from a range of cancer types. Several sensitivity analyses were used to corroborate these findings. Missing data were dealt with using multiple imputations. We attempted to control for possible social desirability of reported smoking behavior by including MCSDS scores in our models. Limitations include the fact that participants were not blinded to their intervention allocation. Most of the participants were women (82.6%); therefore, these results might not be generalizable to men. Self-reported smoking status was not biochemically verified in this study, although self-reports in the general population generally showed good validity, 2 studies among cancer survivors found falsification rates of 48% [48] and 80% [49], whereas a third (substantially smaller study) among thoracic cancer survivors showed relatively good agreement between self-reported and biochemically validated smoking status [50]. Adherence was a limitation in this study and might have influenced the effects in the MyCourse group. The effects were limited in size, and our final sample size was somewhat lower than anticipated. As for any RCT, it remains possible that a true effect was not found in this study (type 2 error). Both the main outcome measure (SC during a single week) and the length of follow-up time (a single year) may have masked the possible long-term impact of sustained cessation on long-term health care use and productivity losses. Now it remains for future studies to evaluate whether sustained SC over several years would have contributed to a reduction in health care and societal costs in the long run.

Conclusions

There was no evidence that at 6 months, the digital MyCourse intervention had a differential effect on cessation rates among cancer survivors compared with a web-based noninteractive information brochure; both conditions led to approximately a quarter of the cancer survivors quitting smoking. The number of cigarettes smoked was reduced by 50% in both groups. At 12 months, assignment to the MyCourse group was associated with a greater reduction in the number of smoked cigarettes at higher costs and higher satisfaction scores compared with the control group. It should be further investigated how to achieve considerably higher quit rates, but this study provides an indication that it is possible to achieve somewhat higher cigarette reduction rates with the help of a digital SC intervention. Although both interventions were low-cost, the noninteractive information brochure was more likely to be economically sustainable.

Acknowledgments

This study was supported by grant TBOS2014–7169 from the Dutch Cancer Society (KWF Kankerbestrijding). The sponsor has no role in the design, data collection, analysis, and interpretation of the data, or in writing the paper or the decision to submit for publication. The authors would like to thank Diede Kesler for her assistance in recruitment and follow-up assessments of the participants and Yvonne Borghans for her assistance in the follow-up assessments of participants.

Conflicts of Interest

The intervention described in this study was developed by the Trimbos Institute (the Netherlands Institute for Mental Health and Addiction).

Multimedia Appendix 1

Additional information on the Methods section.

[[DOCX File, 15 KB - jmir_v24i3e27588_app1.docx](#)]

Multimedia Appendix 2

Attrition and satisfaction with the intervention.

[[DOCX File, 18 KB - jmir_v24i3e27588_app2.docx](#)]

Multimedia Appendix 3

Cost-effectiveness planes and cost-effectiveness acceptability curves after Winsorization.

[[DOCX File, 164 KB - jmir_v24i3e27588_app3.docx](#)]

Multimedia Appendix 4

CONSORT-eHEALTH checklist (V 1.6.1).

[[PDF File \(Adobe PDF File\), 1199 KB - jmir_v24i3e27588_app4.pdf](#)]

References

1. National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. The health consequences of smoking—50 years of progress: a report of the surgeon general. Atlanta, GA: Centers for Disease Control and Prevention; 2014:7-13.
2. Cancer trends progress report: cancer survivors and smoking. National Cancer Institute. 2020. URL: <https://web.archive.org/web/20200515122922/https://progressreport.cancer.gov/after/smoking> [accessed 2020-05-15]
3. Gallaway MS, Glover-Kudon R, Momin B, Puckett M, Lunsford NB, Ragan KR, et al. Smoking cessation attitudes and practices among cancer survivors - United States, 2015. *J Cancer Surviv* 2019;13(1):66-74 [[FREE Full text](#)] [doi: [10.1007/s11764-018-0728-2](https://doi.org/10.1007/s11764-018-0728-2)] [Medline: [30612253](https://pubmed.ncbi.nlm.nih.gov/30612253/)]
4. Ehrenzeller MF, Mayer DK, Goldstein A. Smoking prevalence and management among cancer survivors. *Oncol Nurs Forum* 2018;45(1):55-68. [doi: [10.1188/18.ONF.55-68](https://doi.org/10.1188/18.ONF.55-68)] [Medline: [29251289](https://pubmed.ncbi.nlm.nih.gov/29251289/)]
5. Symes YR, Westmaas JL, Mayer DK, Boynton MH, Ribisl KM, Golden SD. The impact of psychosocial characteristics in predicting smoking cessation in long-term cancer survivors: a time-to-event analysis. *Psychooncology* 2018;27(10):2458-2465. [doi: [10.1002/pon.4851](https://doi.org/10.1002/pon.4851)] [Medline: [30073741](https://pubmed.ncbi.nlm.nih.gov/30073741/)]
6. Cortés-Ibáñez FO, Jaramillo-Calle DA, Vinke PC, Byambasukh O, Corpeleijn E, Sijtsma A, et al. Comparison of health behaviours between cancer survivors and the general population: a cross-sectional analysis of the Lifelines cohort. *J Cancer Surviv* 2020;14(3):377-385 [[FREE Full text](#)] [doi: [10.1007/s11764-020-00854-2](https://doi.org/10.1007/s11764-020-00854-2)] [Medline: [31933151](https://pubmed.ncbi.nlm.nih.gov/31933151/)]
7. Goldstein AO, Ripley-Moffitt CE, Pathman DE, Patsakham KM. Tobacco use treatment at the U.S. National Cancer Institute's designated Cancer Centers. *Nicotine Tob Res* 2013;15(1):52-58 [[FREE Full text](#)] [doi: [10.1093/ntr/nts083](https://doi.org/10.1093/ntr/nts083)] [Medline: [22499079](https://pubmed.ncbi.nlm.nih.gov/22499079/)]
8. Warren GW, Dibaj S, Hutson A, Cummings KM, Dresler C, Marshall JR. Identifying targeted strategies to improve smoking cessation support for cancer patients. *J Thorac Oncol* 2015;10(11):1532-1537 [[FREE Full text](#)] [doi: [10.1097/JTO.0000000000000659](https://doi.org/10.1097/JTO.0000000000000659)] [Medline: [26317914](https://pubmed.ncbi.nlm.nih.gov/26317914/)]
9. Croyle RT, Morgan GD, Fiore MC. Addressing a core gap in cancer care - the NCI moonshot program to help oncology patients stop smoking. *N Engl J Med* 2019;380(6):512-515 [[FREE Full text](#)] [doi: [10.1056/NEJMp1813913](https://doi.org/10.1056/NEJMp1813913)] [Medline: [30601710](https://pubmed.ncbi.nlm.nih.gov/30601710/)]
10. Morgan G, Schnoll RA, Alfano CM, Evans SE, Goldstein A, Ostroff J, et al. National cancer institute conference on treating tobacco dependence at cancer centers. *J Oncol Pract* 2011;7(3):178-182 [[FREE Full text](#)] [doi: [10.1200/JOP.2010.000175](https://doi.org/10.1200/JOP.2010.000175)] [Medline: [21886500](https://pubmed.ncbi.nlm.nih.gov/21886500/)]
11. Derksen JW, Warren GW, Jordan K, Rauh S, Vera García R, O'Mahony D, et al. European practice patterns and barriers to smoking cessation after a cancer diagnosis in the setting of curative versus palliative cancer treatment. *Eur J Cancer* 2020;138:99-108 [[FREE Full text](#)] [doi: [10.1016/j.ejca.2020.07.020](https://doi.org/10.1016/j.ejca.2020.07.020)] [Medline: [32871527](https://pubmed.ncbi.nlm.nih.gov/32871527/)]
12. Farley A, Aveyard P, Kerr A, Naidu B, Dowswell G. Surgical lung cancer patients' views about smoking and support to quit after diagnosis: a qualitative study. *J Cancer Surviv* 2016;10(2):312-319. [doi: [10.1007/s11764-015-0477-4](https://doi.org/10.1007/s11764-015-0477-4)] [Medline: [26298019](https://pubmed.ncbi.nlm.nih.gov/26298019/)]

13. Wells M, Aitchison P, Harris F, Ozakinci G, Radley A, Bauld L, et al. Barriers and facilitators to smoking cessation in a cancer context: a qualitative study of patient, family and professional views. *BMC Cancer* 2017;17(1):348 [FREE Full text] [doi: [10.1186/s12885-017-3344-z](https://doi.org/10.1186/s12885-017-3344-z)] [Medline: [28526000](https://pubmed.ncbi.nlm.nih.gov/28526000/)]
14. Klemp I, Steffenssen M, Bakholdt V, Thygesen T, Sørensen JA. Counseling is effective for smoking cessation in head and neck cancer patients—a systematic review and meta-analysis. *J Oral Maxillofac Surg* 2016;74(8):1687-1694. [doi: [10.1016/j.joms.2016.02.003](https://doi.org/10.1016/j.joms.2016.02.003)] [Medline: [26973223](https://pubmed.ncbi.nlm.nih.gov/26973223/)]
15. Mujcic A, Blankers M, Bommelé J, Boon B, Berman AH, Verdonck-de Leeuw IM, et al. The effectiveness of distance-based interventions for smoking cessation and alcohol moderation among cancer survivors: a meta-analysis. *Psychooncology* 2020;29(1):49-60 [FREE Full text] [doi: [10.1002/pon.5261](https://doi.org/10.1002/pon.5261)] [Medline: [31663182](https://pubmed.ncbi.nlm.nih.gov/31663182/)]
16. Nayan S, Gupta MK, Strychowsky JE, Sommer DD. Smoking cessation interventions and cessation rates in the oncology population: an updated systematic review and meta-analysis. *Otolaryngol Head Neck Surg* 2013;149(2):200-211. [doi: [10.1177/0194599813490886](https://doi.org/10.1177/0194599813490886)] [Medline: [23715685](https://pubmed.ncbi.nlm.nih.gov/23715685/)]
17. Sheeran P, Jones K, Avishai A, Symes YR, Abraham C, Miles E, et al. What works in smoking cessation interventions for cancer survivors? A meta-analysis. *Health Psychol* 2019;38(10):855-865. [doi: [10.1037/hea0000757](https://doi.org/10.1037/hea0000757)] [Medline: [31259596](https://pubmed.ncbi.nlm.nih.gov/31259596/)]
18. Cinciripini PM, Karam-Hage M, Kypriotakis G, Robinson JD, Rabius V, Beneventi D, et al. Association of a comprehensive smoking cessation program with smoking abstinence among patients with cancer. *JAMA Netw Open* 2019;2(9):e1912251 [FREE Full text] [doi: [10.1001/jamanetworkopen.2019.12251](https://doi.org/10.1001/jamanetworkopen.2019.12251)] [Medline: [31560387](https://pubmed.ncbi.nlm.nih.gov/31560387/)]
19. Wild CP, Weiderpass E, Stewart BW. World cancer report: cancer research for cancer prevention. Lyon, France: International Agency for Research on Cancer; 2020:487-498.
20. Bricker JB, Watson NL, Heffner JL, Sullivan B, Mull K, Kwon D, et al. A smartphone app designed to help cancer patients stop smoking: results from a pilot randomized trial on feasibility, acceptability, and effectiveness. *JMIR Form Res* 2020;4(1):e16652 [FREE Full text] [doi: [10.2196/16652](https://doi.org/10.2196/16652)] [Medline: [31951215](https://pubmed.ncbi.nlm.nih.gov/31951215/)]
21. Mujcic A, Blankers M, Boon B, Engels R, van Laar M. Internet-based self-help smoking cessation and alcohol moderation interventions for cancer survivors: a study protocol of two RCTs. *BMC Cancer* 2018;18(1):364 [FREE Full text] [doi: [10.1186/s12885-018-4206-z](https://doi.org/10.1186/s12885-018-4206-z)] [Medline: [29609554](https://pubmed.ncbi.nlm.nih.gov/29609554/)]
22. Lai DT, Cahill K, Qin Y, Tang J. Motivational interviewing for smoking cessation. *Cochrane Database Syst Rev* 2010(1):CD006936. [doi: [10.1002/14651858.CD006936.pub2](https://doi.org/10.1002/14651858.CD006936.pub2)] [Medline: [20091612](https://pubmed.ncbi.nlm.nih.gov/20091612/)]
23. Cviljak M, Stead LF, Hartmann-Boyce J, Sheikh A, Car J. Internet-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2013(7):CD007078. [doi: [10.1002/14651858.CD007078.pub4](https://doi.org/10.1002/14651858.CD007078.pub4)] [Medline: [23839868](https://pubmed.ncbi.nlm.nih.gov/23839868/)]
24. Bricker J, Wyszynski C, Comstock B, Heffner JL. Pilot randomized controlled trial of web-based acceptance and commitment therapy for smoking cessation. *Nicotine Tob Res* 2013;15(10):1756-1764 [FREE Full text] [doi: [10.1093/ntr/ntt056](https://doi.org/10.1093/ntr/ntt056)] [Medline: [23703730](https://pubmed.ncbi.nlm.nih.gov/23703730/)]
25. Sobell LC, Sobell MB. Timeline follow-back: a technique for assessing self-reported ethanol consumption. In: Allen J, Litten RZ, editors. *Measuring alcohol consumption: psychosocial and biological methods*. Totowa, NJ: Humana Press; 1992:41-72.
26. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. *Br J Addict* 1991;86(9):1119-1127. [doi: [10.1111/j.1360-0443.1991.tb01879.x](https://doi.org/10.1111/j.1360-0443.1991.tb01879.x)] [Medline: [1932883](https://pubmed.ncbi.nlm.nih.gov/1932883/)]
27. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption--II. *Addiction* 1993;88(6):791-804. [doi: [10.1111/j.1360-0443.1993.tb02093.x](https://doi.org/10.1111/j.1360-0443.1993.tb02093.x)] [Medline: [8329970](https://pubmed.ncbi.nlm.nih.gov/8329970/)]
28. Crowne DP, Marlowe D. A new scale of social desirability independent of psychopathology. *J Consult Psychol* 1960;24:349-354. [doi: [10.1037/h0047358](https://doi.org/10.1037/h0047358)] [Medline: [13813058](https://pubmed.ncbi.nlm.nih.gov/13813058/)]
29. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Ann Med* 2001;33(5):337-343. [doi: [10.3109/07853890109002087](https://doi.org/10.3109/07853890109002087)] [Medline: [11491192](https://pubmed.ncbi.nlm.nih.gov/11491192/)]
30. Ware Jr JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30(6):473-483. [Medline: [1593914](https://pubmed.ncbi.nlm.nih.gov/1593914/)]
31. Brazier JE, Roberts J. The estimation of a preference-based measure of health from the SF-12. *Med Care* 2004;42(9):851-859. [doi: [10.1097/01.mlr.0000135827.18610.0d](https://doi.org/10.1097/01.mlr.0000135827.18610.0d)] [Medline: [15319610](https://pubmed.ncbi.nlm.nih.gov/15319610/)]
32. Kriz D, Nübling R, Steffanowski A, Wittmann WW, Schmidt J. Patientenzufriedenheit in der stationären rehabilitation: psychometrische reanalyse des ZUF-8 auf der basis multizentrischer stichproben verschiedener indikation. *Z Med Psychol* 2008;17(2-3):67-79 [FREE Full text]
33. Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Zorginstituut Nederland. 2016. URL: <https://www.zorginstituutnederland.nl/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg> [accessed 2019-01-11]
34. Medicijnkosten. Zorginstituut Nederland. 2019. URL: <https://www.medicijnkosten.nl/> [accessed 2019-11-01]
35. van Roijen L, van Straten A, Tiemens B, Donker MC. Manual trimbos/iMTA questionnaire for costs associated with psychiatric illness (TIC-P). Institute for Medical Technology Assessment 2002;2 [FREE Full text]

36. Duffy SA, Ronis DL, Valenstein M, Lambert MT, Fowler KE, Gregory L, et al. A tailored smoking, alcohol, and depression intervention for head and neck cancer patients. *Cancer Epidemiol Biomarkers Prev* 2006;15(11):2203-2208 [[FREE Full text](#)] [doi: [10.1158/1055-9965.EPI-05-0880](https://doi.org/10.1158/1055-9965.EPI-05-0880)] [Medline: [17119047](https://pubmed.ncbi.nlm.nih.gov/17119047/)]
37. van Buuren S, 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](https://doi.org/10.18637/jss.v045.i03)]
38. Atkins DC, Baldwin SA, Zheng C, Gallop RJ, Neighbors C. A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychol Addict Behav* 2013;27(1):166-177 [[FREE Full text](#)] [doi: [10.1037/a0029508](https://doi.org/10.1037/a0029508)] [Medline: [22905895](https://pubmed.ncbi.nlm.nih.gov/22905895/)]
39. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. Oxford, UK: Oxford University Press; 2015.
40. Husereau D, Drummond M, Petrou S, Carswell C, Moher D, Greenberg D, CHEERS Task Force. Consolidated health economic evaluation reporting standards (CHEERS) statement. *Cost Eff Resour Alloc* 2013;11(1):6 [[FREE Full text](#)] [doi: [10.1186/1478-7547-11-6](https://doi.org/10.1186/1478-7547-11-6)] [Medline: [23531194](https://pubmed.ncbi.nlm.nih.gov/23531194/)]
41. Lamers LM, Stalmeier PF, McDonnell J, Krabbe PF, van Busschbach JJ. Measuring the quality of life in economic evaluations: the Dutch EQ-5D tariff. *Ned Tijdschr Geneesk* 2005;149(28):1574-1578. [Medline: [16038162](https://pubmed.ncbi.nlm.nih.gov/16038162/)]
42. Graham AL, Carpenter KM, Cha S, Cole S, Jacobs MA, 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](https://doi.org/10.2147/SAR.S101660)] [Medline: [27274333](https://pubmed.ncbi.nlm.nih.gov/27274333/)]
43. Taylor GM, Dalili MN, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. *Cochrane Database Syst Rev* 2017;9(9):CD007078 [[FREE Full text](#)] [doi: [10.1002/14651858.CD007078.pub5](https://doi.org/10.1002/14651858.CD007078.pub5)] [Medline: [28869775](https://pubmed.ncbi.nlm.nih.gov/28869775/)]
44. Emmons KM, Puleo E, Sprunck-Harrild K, Ford J, Ostroff JS, Hodgson D, et al. Partnership for health-2, a web-based versus print smoking cessation intervention for childhood and young adult cancer survivors: randomized comparative effectiveness study. *J Med Internet Res* 2013;15(11):e218 [[FREE Full text](#)] [doi: [10.2196/jmir.2533](https://doi.org/10.2196/jmir.2533)] [Medline: [24195867](https://pubmed.ncbi.nlm.nih.gov/24195867/)]
45. Götze H, Taubenheim S, Dietz A, Lordick F, Mehnert A. Comorbid conditions and health-related quality of life in long-term cancer survivors-associations with demographic and medical characteristics. *J Cancer Surviv* 2018;12(5):712-720. [doi: [10.1007/s11764-018-0708-6](https://doi.org/10.1007/s11764-018-0708-6)] [Medline: [30097854](https://pubmed.ncbi.nlm.nih.gov/30097854/)]
46. Hoogendoorn M, Feenstra TL, Hoogenveen RT, Rutten-van Mölken MP. Long-term effectiveness and cost-effectiveness of smoking cessation interventions in patients with COPD. *Thorax* 2010;65(8):711-718. [doi: [10.1136/thx.2009.131631](https://doi.org/10.1136/thx.2009.131631)] [Medline: [20685746](https://pubmed.ncbi.nlm.nih.gov/20685746/)]
47. Graham AL, Chang Y, Fang Y, Cobb NK, Tinkelman DS, Niaura RS, et al. Cost-effectiveness of internet and telephone treatment for smoking cessation: an economic evaluation of The iQUITT study. *Tob Control* 2013;22(6):e11 [[FREE Full text](#)] [doi: [10.1136/tobaccocontrol-2012-050465](https://doi.org/10.1136/tobaccocontrol-2012-050465)] [Medline: [23010696](https://pubmed.ncbi.nlm.nih.gov/23010696/)]
48. Klesges RC, Krukowski RA, Klosky JL, Liu W, Srivastava DK, Boyett JM, et al. Efficacy of a tobacco quitline among adult cancer survivors. *Prev Med* 2015;73:22-27 [[FREE Full text](#)] [doi: [10.1016/j.ypmed.2014.12.019](https://doi.org/10.1016/j.ypmed.2014.12.019)] [Medline: [25572620](https://pubmed.ncbi.nlm.nih.gov/25572620/)]
49. Klesges RC, Krukowski RA, Klosky JL, Liu W, Srivastava DK, Boyett JM, et al. Efficacy of a tobacco quitline among adult survivors of childhood cancer. *Nicotine Tob Res* 2015;17(6):710-718 [[FREE Full text](#)] [doi: [10.1093/ntr/ntu216](https://doi.org/10.1093/ntr/ntu216)] [Medline: [25335944](https://pubmed.ncbi.nlm.nih.gov/25335944/)]
50. Park ER, Japuntich S, Temel J, Lanuti M, Pandiscio J, Hilgenberg J, et al. A smoking cessation intervention for thoracic surgery and oncology clinics: a pilot trial. *J Thorac Oncol* 2011;6(6):1059-1065 [[FREE Full text](#)] [doi: [10.1097/JTO.0b013e318215a4dc](https://doi.org/10.1097/JTO.0b013e318215a4dc)] [Medline: [21512406](https://pubmed.ncbi.nlm.nih.gov/21512406/)]

Abbreviations

AUDIT: Alcohol Use Disorders Identification Test

CEAC: cost-effectiveness acceptability curve

CONSORT: Consolidated Standards of Reporting Trials

EQ-5D-5L: 5-level EuroQoL 5

FTND: Fagerström Test for Nicotine Dependence

GLMM: Generalized Linear Mixed Model

ICER: incremental cost-effectiveness ratio

IRR: incidence rate ratio

MCSDS: Marlowe-Crowne Social Desirability Scale

OR: odds ratio

QALY: quality-adjusted life year

RCT: randomized controlled trial

RR: relative risk

SC: smoking cessation

TLFB: timeline Followback

ZUF-8: Fragebogen zur Messung der Patientenzufriedenheit (Patient Satisfaction Questionnaire)

Edited by T Leung; submitted 29.01.21; peer-reviewed by J Li, A Munro; comments to author 06.03.21; revised version received 28.04.21; accepted 23.12.21; published 17.03.22.

Please cite as:

Mujcic A, Blankers M, Boon B, Verdonck-de Leeuw IM, Smit F, van Laar M, Engels R

Effectiveness, Cost-effectiveness, and Cost-Utility of a Digital Smoking Cessation Intervention for Cancer Survivors: Health Economic Evaluation and Outcomes of a Pragmatic Randomized Controlled Trial

J Med Internet Res 2022;24(3):e27588

URL: <https://www.jmir.org/2022/3/e27588>

doi: [10.2196/27588](https://doi.org/10.2196/27588)

PMID: [35297777](https://pubmed.ncbi.nlm.nih.gov/35297777/)

©Ajla Mujcic, Matthijs Blankers, Brigitte Boon, Irma M Verdonck-de Leeuw, Filip Smit, Margriet van Laar, Rutger Engels. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 17.03.2022. 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.

Review

Digital Education for Health Professionals: An Evidence Map, Conceptual Framework, and Research Agenda

Lorraine Tudor Car^{1,2}, MSc, MD, PhD; Selina Poon³, BSc, MSc, PhD; Bhone Myint Kyaw³, MBBS, MSc, PhD; David A Cook⁴, MD, MHPE; Victoria Ward⁵, PhD; Rifat Atun⁶, MBBS, FRCGP, FRCP; Azeem Majeed², PhD; Jamie Johnston⁷, PhD; Rianne M J J van der Kleij⁸, PhD; Mariam Molokhia⁹, PhD; Florian V Wangenheim¹⁰, PhD; Martin Lupton^{11,12}, MBBS, MRCOG; Niels Chavannes⁸, MD, PhD; Onyema Ajuebor¹³, MBBS, MSc; Charles G Prober⁷, MD; Josip Car^{2,3}, MD, PhD

¹Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

²Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom

³Centre for Population Health Sciences, Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

⁴Office of Applied Scholarship and Education Science, School of Continuous Professional Development, Mayo Clinic College of Medicine and Science, Division of General Internal Medicine, Mayo Clinic, Rochester, MN, United States

⁵Department of Pediatrics, Lucile Packard Children's Hospital, Stanford, CA, United States

⁶Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Harvard, Boston, MA, United States

⁷Stanford Center for Health Education's Digital MedIC Initiative, Stanford School of Medicine, Stanford University, Stanford, CA, United States

⁸Department of Public Health and Primary Care, Leiden University Medical Centre, Leiden, Netherlands

⁹School of Population Health and Environmental Sciences, Faculty of Life Sciences and Medicine, King's College London, London, United Kingdom

¹⁰Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

¹¹Faculty of Medicine, Imperial College London, London, United Kingdom

¹²The Chelsea and Westminster Hospital, Chelsea, London, United Kingdom

¹³Health Workforce Department, World Health Organization, Geneva, Switzerland

Corresponding Author:

Josip Car, MD, PhD

Centre for Population Health Sciences

Lee Kong Chian School of Medicine

Nanyang Technological University

11 Mandalay Road, Clinical Science Building, Novena

Singapore, 308232

Singapore

Phone: 65 69041258

Email: josip.car@ntu.edu.sg

Abstract

Background: Health professions education has undergone major changes with the advent and adoption of digital technologies worldwide.

Objective: This study aims to map the existing evidence and identify gaps and research priorities to enable robust and relevant research in digital health professions education.

Methods: We searched for systematic reviews on the digital education of practicing and student health care professionals. We searched MEDLINE, Embase, Cochrane Library, Educational Research Information Center, CINAHL, and gray literature sources from January 2014 to July 2020. A total of 2 authors independently screened the studies, extracted the data, and synthesized the findings. We outlined the key characteristics of the included reviews, the quality of the evidence they synthesized, and recommendations for future research. We mapped the empirical findings and research recommendations against the newly developed conceptual framework.

Results: We identified 77 eligible systematic reviews. All of them included experimental studies and evaluated the effectiveness of digital education interventions in different health care disciplines or different digital education modalities. Most reviews included studies on various digital education modalities (22/77, 29%), virtual reality (19/77, 25%), and online education (10/77,

13%). Most reviews focused on health professions education in general (36/77, 47%), surgery (13/77, 17%), and nursing (11/77, 14%). The reviews mainly assessed participants' skills (51/77, 66%) and knowledge (49/77, 64%) and included data from high-income countries (53/77, 69%). Our novel conceptual framework of digital health professions education comprises 6 key domains (context, infrastructure, education, learners, research, and quality improvement) and 16 subdomains. Finally, we identified 61 unique questions for future research in these reviews; these mapped to framework domains of education (29/61, 47% recommendations), context (17/61, 28% recommendations), infrastructure (9/61, 15% recommendations), learners (3/61, 5% recommendations), and research (3/61, 5% recommendations).

Conclusions: We identified a large number of research questions regarding digital education, which collectively reflect a diverse and comprehensive research agenda. Our conceptual framework will help educators and researchers plan, develop, and study digital education. More evidence from low- and middle-income countries is needed.

(*J Med Internet Res* 2022;24(3):e31977) doi:[10.2196/31977](https://doi.org/10.2196/31977)

KEYWORDS

digital education; health professions education; evidence map; systematic review; research questions; conceptual framework; mobile phone

Introduction

The world is faced with a shortage and an unequal distribution of health workforces across low-income, middle-income, and high-income countries [1]. The shortfalls and inequitable distributions affect the likelihood of reaching the United Nations' third Sustainable Development Goal—health and well-being for all by 2030 [1-5]. To enable an increase in and a more equitable distribution of competent health workforce, there is a need for more effective and accessible health professions education.

The use of digital technology in health professions education can help in overcoming some of the health workforce-related challenges by providing more accessible, standardized, relevant, timely, and affordable medical education and training [6,7]. Until recently, digital education was perceived as primarily supporting in-person health professions education [8]. The social distancing measures introduced to control the COVID-19 pandemic have dramatically changed the delivery of health professions education worldwide. Many medical schools and health professions education institutions had to pivot to digital education [9,10]. With this sudden shift, research and evidence in digital health professions education have become even more important.

The evidence on digital education has grown substantially in recent years and has been the subject of many systematic reviews. Existing reviews seem to mostly focus on the effectiveness of different digital education modalities [11-16]. However, the adoption of digital education is complex and includes other research questions, in addition to its effectiveness. It is important to identify evidence that already exists and evidence gaps across the full scope of relevant questions to inform and guide future research and reduce research waste. To address this need, we seek to (1) create a map of existing research, (2) develop a conceptual framework outlining key components of digital education, and (3) highlight specific research questions across a comprehensive research framework. We do this by systematically identifying and analyzing previous systematic reviews on digital health education.

Methods

We used an evidence map methodology to identify and summarize systematic reviews on digital health professions education [17]. We also developed a novel conceptual framework using an established methodological approach [18] and identified specific research questions in alignment with this conceptual framework.

Study Identification

To identify relevant systematic reviews on different types of digital health education for health professionals, we used a comprehensive search strategy mentioned in [Multimedia Appendix 1](#), including key terms for participants (eg, *health professionals*, *health personnel*, and *students*), intervention (eg, *e-learning*, *patient simulation*, and *serious games*), and article type (eg, *systematic reviews* and *evidence synthesis*). We searched the following major bibliographic databases for studies published between January 1, 2014, and July 21, 2020, without any restrictions on language or study design: MEDLINE (Ovid), Embase (Ovid), Cochrane Library, Educational Research Information Center (EBSCO), and CINAHL (EBSCO). We also manually checked the reference lists of the included systematic reviews for other potentially relevant systematic reviews on digital education for health professionals. In addition, we searched Google, Google Scholar, ResearchGate, and OpenGrey for any other studies that might be relevant for our topic, using keywords *systematic review*, *digital education*, *health professionals*, *health professions education*, *eLearning*, and *e-learning*, and reviewing either the first 10 pages or 500 results.

Eligibility Criteria

We included systematic reviews focusing on digital education for health care professionals in preservice (ie, student) and in-service (ie, postdegree, including postgraduate trainees and those in independent practice) positions [19]. This includes disciplines such as medicine, dentistry, nursing and midwifery, medical diagnostic and treatment technology, physiotherapy and rehabilitation, and pharmacy. Practitioners of traditional, alternative, and complementary medicine were excluded. Digital health professions education refers to health professions education that is conducted using digital technology [20] and

includes modalities ranging from the basic conversion of content into a digital format (eg, a book converted into a PDF or HTML format) to more complex applications such as mobile education, digital games, virtual patients, and virtual reality (VR). Systematic reviews were included if they focused on ≥ 1 modality of digital education (as defined in [Textbox 1](#)) delivered in a stand-alone or blended format [20]. We defined blended education as education that incorporates aspects of traditional

and digital education. Traditional education was defined as education that encompasses the use of nondigital educational materials (eg, textbooks or models) or in-person human interactions. We included systematic reviews of all the types of studies. We excluded older reviews because of the rapid evolution of the field, with the assumption that most of the active research questions from the reviews published >5 years ago would be collated in more recent reviews.

Textbox 1. Digital education technologies and modalities and working definitions and descriptions.

Offline digital education

- Education delivery requires no internet or local area network connection and can be delivered through external media, including CD-ROM, external hard disk, and USB stick [21].

Online digital education

- Computer-assisted instruction using the internet or a local intranet as the means of delivery, also referred to as online, internet-based, or networked [22,23], includes multiple media formats (eg, text, videos, and images and online discussion (eg, via email, chat, or videoconferencing) and is designed to be primarily delivered on PCs.

Massive open online course

- A (free) online course available over the internet to a large number of geographically dispersed participants [24]

Mobile education (m-Learning)

- Flexible and accessible learning delivered via personal mobile devices, such as smartphones and tablets [25]

Serious gaming and gamification

- Knowledge and training activities are set within a competitive activity. Games are intended to promote the development of knowledge, cognitive skills, or psychomotor skills in a virtual environment [26].

Virtual reality

- Interactive exploration of a digital (3D) multimedia environment can reflect a real-world environment (eg, clinic) or an artificial or surreal context (eg, positioning the learner within the human body) [16,27].

Virtual patient

- A computer program that simulates real-life clinical scenarios where students take on the role of a health professional and obtain a patient's history, conduct a physical examination, and make diagnostic and therapeutic decisions [28]

High-fidelity manikins

- Realistic, computerized mannequins that mimic elements of human physiology (eg, breathing and heart rhythm) and are used to simulate a real-life clinical scenario [29].

Blended education

- The use of digital education modalities in combination with traditional education methods

Traditional education

- Education that uses nondigital educational material (eg, textbook or model) or in-person human interaction

Study Selection

The search results from different databases were combined in a single EndNote library, and duplicate citations were removed. A total of 2 review authors (SP and BMK) screened all titles and abstracts for inclusion independently and in duplicate. Disagreements during the title and abstract screening were resolved by consensus. Full texts of articles considered eligible or uncertain based on the title and abstract screening were

retrieved and screened independently and in duplicate by the same 2 authors.

Data Extraction

From the included systematic reviews, 2 authors (SP and BMK) used a standardized form to independently extract information on the review aim; the study design, participants, interventions, comparators, and outcomes of the original research studies included; the method used to appraise the quality of the included studies (eg, risk of bias) or overall evidence (eg, the Grading

of Recommendations Assessment, Development and Evaluation assessment); and recommendations for future research. We classified the outcomes according to the definitions presented in [Multimedia Appendix 2](#). We also extracted information on all additional outcomes reported in the included reviews that did not correspond to our predefined outcome-related definitions. We classified the systematic review in terms of the single digital modality they most focused on, according to our framework ([Textbox 1](#)). In most instances, it was clear that a given review focused predominantly on 1 modality. Less often, a review encompassed multiple modalities equally, in which case we classified it as digital education; that is, the use of digital technology in health professions education in general. Finally, we identified recommendations for future research by extracting exact quotes from each review that articulated such recommendations. At every stage, disagreements between the review authors were resolved through discussion and input from the third author (LTC).

Analysis

Authors SP and BMK rephrased the quoted research recommendations into research questions and then refined these by applying consistent terminology, removing duplicates, and merging or subdividing themes. Questions that focused on specific digital modalities (eg, online modules) were rewritten to make them relevant to the broader research agenda for digital health education. The final list of research questions was refined by LTC.

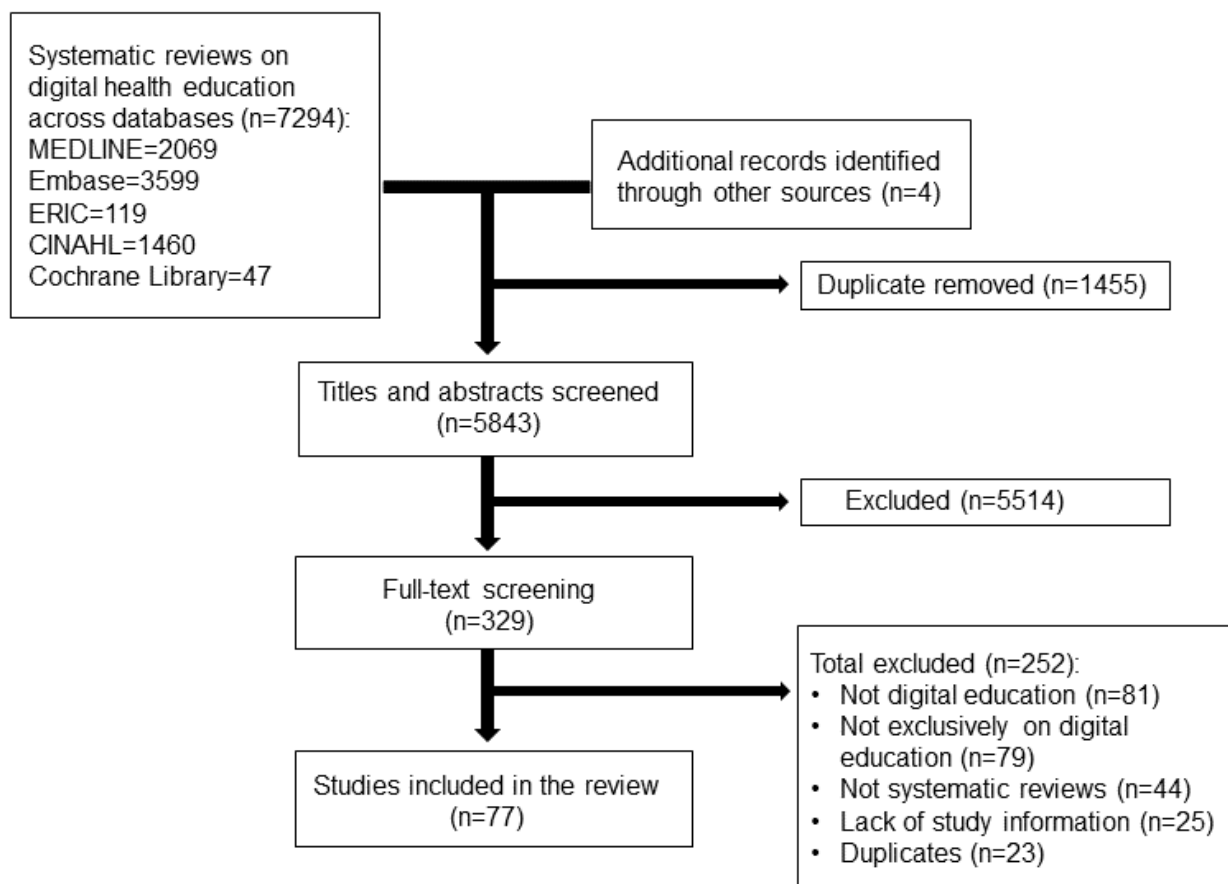
We developed a conceptual framework outlining various digital health professions’ education components according to the methodology described by Jabareen [18]. We consulted and built on our previous conceptual work in this area and existing frameworks for the implementation or adoption of digital education generally [30-39]. We identified these frameworks through a focused literature search on PubMed, Google, and Google Scholar. On the basis of the discussions and consensus among the review authors, key domains and subdomains were finalized. The framework and its components were represented diagrammatically. Recommendations for future research were classified according to the proposed framework in parallel by 2 authors (SP and BMK). Discrepancies were resolved through consensus and with the guidance of the third author (LTC). On the basis of the analysis of the included reviews and the observed gaps in the literature, we outline a research agenda for digital health professions education and present it in the *Results* section.

Results

Search Results

Of 7294 systematic reviews from our initial search, we identified 73 (1%) eligible systematic reviews ([Figure 1](#)). Another 4 systematic reviews were identified through Google, Google Scholar, ResearchGate, and OpenGrey. In total, 77 systematic reviews were included for data extraction.

Figure 1. Study flow diagram. ERIC: Educational Research Information Center.



The number of published systematic reviews increased over time, from 6% (5/77) published in 2014 to 56% (43/77) published in 2019. The systematic reviews focused on digital education for health professions students (17/77, 22% studies), postgraduate or independently practicing (18/77, 23% studies) health professionals, or both (42/77, 55% studies). Most of the systematic reviews were on digital education in surgery (13/77, 17% studies) or health professions education in general (ie, those not focusing on a particular type of practitioner, health care area, or topic; 36/77, 47% studies), followed by nursing (11/77, 14% studies; [Table 1](#); [Multimedia Appendix 3](#) [11-16,32,34,40-109]).

A breakdown of the digital modalities being investigated in the included systematic reviews is shown in [Figure 2](#). Most systematic reviews focused on digital education in general (22/77, 29%), VR (19/77, 15%), and online education (10/77, 13%). Of the 19 reviews on VR, 17 (89%) were on VR complemented with physical objects or devices such as probes or handles and focused on psychomotor, procedural, or technical skill development. There were fewer reviews published on m-Learning (6/77, 8%), digital game-based learning (3/77, 4%), and virtual patients (2/77, 3%).

The most common outcomes in the included reviews were health professionals' knowledge (49/77, 64%), skills (51/77, 66%), attitudes about the clinical topic (13/77, 17%), and satisfaction with digital education (18/77, 23%). Most systematic reviews

compared the effectiveness of digital education to traditional education (ie, nondigital; 59/77, 77%) or other forms of digital education (35/77, 45%; [Table 1](#)). Most reviews reported only immediate, short-term outcomes; 22% (17/77) of reviews reported the impact of digital education on long-term delayed outcomes; that is, outcomes assessed with delay after the intervention [34,40-55]. Most reviews appraised methods using the Risk of Bias tool [110] only (24/77, 31%), followed by Grading of Recommendations, Assessment, Development, and Evaluation [111] (which also includes the risk of bias assessment; 22/77, 29%) and Medical Education Research Study Quality Instrument [112] (10/77, 13%). Of the 77 studies, 9 (12%) reviews did not report on the quality appraisal of the included evidence, whereas the remaining 14 (18%) reviews used different instruments to determine the evidence quality ([Table 1](#); [Multimedia Appendix 3](#) [11-16,32,34,40-107,109]). The included reviews mostly reported original research to be low or very low quality of evidence or reported unclear or high risk of bias in most studies ([Multimedia Appendix 3](#) [11-16,32,34,40-107,109]). Most systematic reviews reported data from high-income countries (14/77, 18% systematic reviews) or both middle- and high-income countries (42/77, 55% systematic reviews). Only 4% (3/77) of the systematic reviews included studies from low-income countries [11,56,57]. Approximately 29% (22/77) of the included reviews did not report the setting of the included studies.

Table 1. Characteristics of the included systematic reviews (N=77).

Characteristics of the systematic reviews and the evidence they include	Studies, n (%)
Type of participants	
Medical students	5 (6)
Medical students and physicians	9 (12)
Physicians	17 (22)
Dentistry students	3 (4)
Dentistry students and dentists	2 (3)
Nursing students	8 (10)
Nursing students and nurses	3 (4)
Mixed students	2 (3)
Mixed students and HCPs ^a	19 (25)
Mixed HCPs	9 (12)
Level of education	
Postdegree: practicing HCPs	10 (13)
Postdegree: trainees ^b	5 (6)
Postdegree: mix of trainees ^b and practicing HCPs	3 (4)
Student	17 (22)
Mixed student and postdegree	42 (55)
Clinical topics	
General health professions education	23 (30)
Surgery	14 (18)
Nursing	8 (10)
Life support or trauma management (resuscitation skills)	3 (4)
Radiology	7 (9)
Endoscopy	3 (4)
Other	19 (25)
Setting	
High-income countries only	26 (34)
High-income and middle-income countries	45 (58)
High-, middle-, and low-income countries	4 (5)
Middle- and low-income countries	1 (1)
Information not available	1 (1)
Modality	
Digital education	22 (29)
Virtual reality	19 (25)
Online	10 (13)
Offline	6 (8)
Mobile learning	6 (8)
High-fidelity manikins	5 (6)
Online and offline	4 (5)
Digital serious games	3 (4)
Virtual patient	2 (3)

Characteristics of the systematic reviews and the evidence they include	Studies, n (%)
Comparison	
No intervention	25 (32)
Traditional education	56 (73)
Digital intervention	35 (45)
Other	6 (8)
Quality appraisal	
Risk of bias	24 (31)
Grading of Recommendations, Assessment, Development, and Evaluations	22 (29)
Medical Education Research Study Quality Instrument	10 (13)
Best Evidence in Medical Education reviews	2 (3)
The Jadad scale	2 (3)
Joanna Briggs Institute Meta-Analysis of Statistics Assessment and Review Instrument	2 (3)
Methodological index for non-randomized studies	1 (1)
Newcastle–Ottawa Scale	1 (1)
Other	5 (6)
Not reported	8 (10)
Outcomes	
Knowledge	49 (64)
Skills	51 (66)
Satisfaction	18 (23)
Patient outcomes	20 (26)
Performance ^c	19 (25)
Attitude	13 (17)
Behavioral	8 (10)
Number of included studies in the review	
<10	24 (31)
10-19	27 (35)
20-29	10 (13)
30-39	8 (10)
≥40	8 (10)
Study designs included in the review	
Randomized controlled trials	68 (88)
Other experimental studies ^d	8 (10)
Cross-sectional studies	5 (6)
Qualitative studies	3 (4)
Pre-post studies	12 (16)
Cohort studies	8 (10)
Other or mixed ^e	34 (44)
The conceptual framework domain or subdomain the reviews focus on	
Education—design	77 (100)
Education—content	77 (100)
Education—evaluation	9 (12)

Characteristics of the systematic reviews and the evidence they include	Studies, n (%)
Education—pedagogy	5 (6)
Education—engagement	3 (4)
Context—settings	1 (1)

^aHCP: health care professional.

^bIncludes residents, novices, trainees, and fellows.

^cDefined in the included systematic reviews as a combination of skills and behavioral changes as a result of the intervention.

^dIncludes quasi-randomized controlled trials, nonrandomized controlled trials, before-and-after studies, and interrupted time series designs.

^eIncludes study designs not described above or a combination of different study designs.

Figure 2. The number of systematic reviews on different digital modalities according to the year of publication. m-Learning: mobile learning.



Conceptual Framework of Digital Health Professions Education

To outline different aspects of digital health professions education and identify gaps in the literature, we developed a novel conceptual framework (Figure 3) grounded in key findings of these systematic reviews together with 7 existing frameworks for digital education general [35-39,113-117] and a framework we developed previously for health professions education [33].

Broadly, the fundamental domains include an enabling and supportive *context*, sound *infrastructure*, and the optimal use of *education* tools and processes. The context is a combination of institutional norms, sociocultural norms, and settings in which the learner resides, as well as the level of education the learner is at. Subdomains of the context have a direct impact on the

infrastructure components required and available for the delivery of digital education—digital and physical spaces, policies and regulatory standards, and human resources. Both context and infrastructure components are important in consideration of health professions education. *Learners*, individually and as part of a larger group, are at the core of digital health education, and their needs, preferences, prior expertise, and competencies should shape how education is delivered. The interaction among components within and across each layer is dynamic, with different parts being interconnected, as reflected using dotted lines to separate context, infrastructure, education, and learners. Studying and identifying optimal relationships between the components are handled by the *research* and *quality assurance* blocks, which are connected to the rest of the framework. Table 2 provides the detailed operational definitions for each domain of the framework.

Figure 3. Conceptual framework of digital health education for healthcare professionals. CME: continuing medical education; CPD: continuing professional development; IT: information technology.

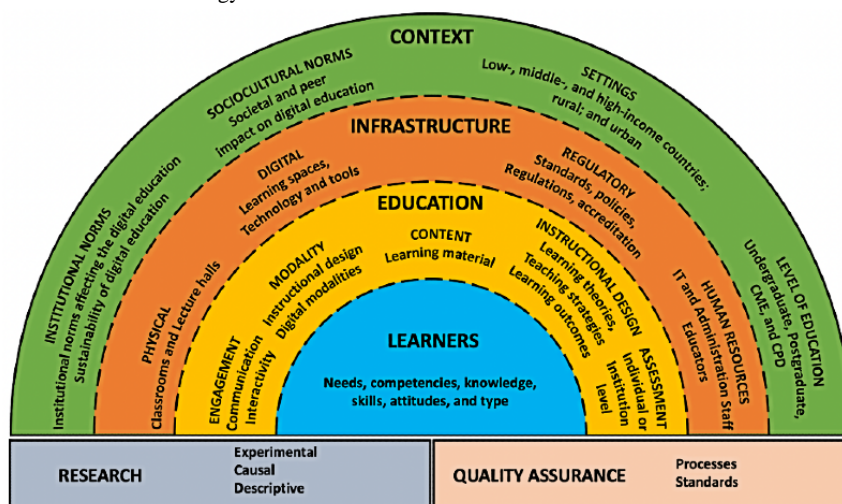


Table 2. Definitions of digital health professions education conceptual framework components.

Domain and subdomain	Definition
Context	
Sociocultural norms	The acceptability and adoption of digital education as a form and norm of education within the society
Institutional norms	The acceptability, impact, considerations, and processes concerning the adoption of digital education at the institutional level
Settings	The setting in which digital health education is conducted or implemented, including clinical or classroom environments; low-, middle-, and high-income countries; and rural or urban environments
Level of education	The impact and integration of digital education with other forms of education (eg, inter- and intraprofessional training opportunities) and clinical work in which participants are engaged
Infrastructure	
Physical	The physical learning space within which the in-person component of blended digital health education is taking place
Digital	The information and communication technology devices (both hardware and software) to support and create learning environments (virtual environments, digital networks, technological modifications) or media for digital health education, as well as the speed and capacity of internet access
Regulatory	Policies and regulatory standards for health professionals' licensing and accreditation, as well as those relating to the design and delivery of digital health professions education
Human resources	The human resources required for digital health education to be maintained and sustained, including educators, administrators, and information technology staff
Education	
Modality	The choice and configuration of digital education modality (eg, online learning and m-Learning) and its potential blending with in-person education
Instructional design	The method and practice of digital health professions education encompassing teaching strategies, learning principles, learning outcomes, and the assessment approach
Content	Health professions education area, discipline, theme, or topic delivered via digital education
Engagement	The level of communication, interactivity, or immersion of participants taking part in digital health professions education
Assessment	Measurement of digital health professions education conducted at the individual and institutional level to determine its impact on educational and clinical outcomes
Learners	Health professionals with distinctive needs, competencies, digital literacy, knowledge, skills, and attitudes toward working and learning, both individually or as a group
Research	Systematic study of digital health professions education to create and disseminate new knowledge and allow for more effective and efficient adoption, implementation, and transfer of interventions to various contexts—this encompasses experimental, observational, descriptive, and qualitative research
Quality assurance	A context-specific and systematic evaluation of practices and procedures to understand the current state and improve the performance of digital health education in a particular setting

Research Questions From the Included Systematic Reviews

We identified 318 discrete research questions posed in these 77 articles, from which we distilled a final list of 61 (19.2%) distinct questions covering 14 of the 16 subdomains of the above framework (Table 2; Multimedia Appendix 4 [11-16,32,34,40-95,98-107]). Research questions that spanned multiple subdomains were assigned by investigator consensus to the most relevant single subdomain. None of the included systematic reviews posed questions primarily directed at the *physical* infrastructure or *quality assurance* subdomains. We identified 26% (16/61) of questions relating to context, classified into four subdomains: *sociocultural norms*, *institutional norms*, *settings*, and *level of education*. Approximately 15% (9/61) of research questions (3 per subdomain) were identified for the *digital*, *regulatory*, and *human resources* subdomains within the infrastructure domain. Most of the research questions, 48% (29/61) approximately, were categorized in the education domain, which encompasses *modality*, *instructional design*,

content, *engagement*, and *evaluation* subdomains. Approximately 5% (3/61) of *research* questions each were categorized in the *learners* and *research* domains.

Classifying Research Questions Addressed by Existing Systematic Reviews

We also classified the included systematic reviews based on their research questions and using our conceptual framework (Table 3; Multimedia Appendix 3 [11-16,32,34,40-107,109]). The research questions addressed by existing systematic reviews mostly revolved around digital education *modality* (ie, the effectiveness of various digital education modalities delivered as stand-alone or blended interventions) and *content* (ie, the effectiveness of digital education within a particular health care area or discipline). Some reviews assessed interactivity (*engagement*), various aspects of *instructional design* in digital education, the impact of digital education on institutional outcomes (*context—institutional norms*), and the impact of *setting* (eg, low-income and middle-income countries) on learning outcomes.

Table 3. Research questions identified from the included systematic reviews on health professions digital education.

Research questions identified from included systematic reviews	Conceptual framework domain (subdomain)	Systematic reviews' references
How do cost and cost-related outcomes influence the adoption of digital technology in health professions education?	Context (sociocultural norms)	[41,56,58-65]
How can policy makers be organized to adopt digital education as part of health professions education?	Context (sociocultural norms)	[56,66]
How do cultural factors within different countries determine the use of digital education for health professions training?	Context (sociocultural norms)	[66]
How does providing access to digital education improve the learning outcomes of health professionals?	Context (sociocultural norms)	[14,40,41,43,46,53,56,57,66-77]
What is the long-term cost-effectiveness of digital education compared with traditional education for health professionals?	Context (institutional norms)	[12-14,16,47,61,78]
How does health professions' digital education affect individual and health services outcomes and organizational practice?	Context (institutional norms)	[11-14, 16, 32, 34, 40, 42-45, 48, 49, 51-53, 56, 60, 62, 64, 67-70, 75, 79-88]
Is health professions' digital education more time efficient than traditional education?	Context (institutional norms)	[46]
What is the feasibility of implementing digital technology for health professions education in different socioeconomic settings?	Context (setting)	[13,14,16,43,56,57,78,89,90]
What are the short- and long-term effects of using digital technology for health professions education in different socioeconomic settings?	Context (setting)	[32,43,47,50,54,60,78,82,89-91]
Is digital education for health professionals effective in different socioeconomic settings?	Context (setting)	[13,40,91,92]
What are the resource requirements to implement digital education in different socioeconomic settings?	Context (setting)	[85,93]
What are the challenges of setting up digital education for health professionals training in different socioeconomic settings?	Context (setting)	[85]
What is the differential impact of digital education on the clinical performance of trainee or expert surgeons?	Context (level)	[94]
How can digital education for health professionals be integrated into normal work practices?	Context (level)	[68]
How can digital technology be incorporated into current health professions' education and training curriculum to improve learning outcomes?	Context (level)	[42,46,47,54,61,62,78,90]
Is digital education effective in improving health professionals' knowledge and skills performance in the clinical setting?	Context (level)	[11-14, 16, 32, 34, 40, 42-45, 48, 49, 51-53, 56, 60, 62, 64, 67-70, 75, 79-88]
Which features of digital education (eg, technical features, fidelity, safety, and adaptability) affect the learning outcomes of health professions education?	Infrastructure (digital)	[13,95]
What are the minimum requirements for the digital technology used to achieve the effectiveness of digital health professions education?	Infrastructure (digital)	[85]
What are the technical resources needed to deliver digital education to health care professionals?	Infrastructure (digital)	[61]
How should educators delivering digital health education be assessed and accredited?	Infrastructure (regulatory)	[47]
What are the best practices for the development, evaluation, and use of digital health education in health professions education?	Infrastructure (regulatory)	[14]
Is the use of accreditation-related milestones in digital health education effective?	Infrastructure (regulatory)	[78]
What digital skills should instructors facilitating digital health education be competent in?	Infrastructure (human resources)	[47]
How does the digital competence of teachers affect health professions learning outcomes from digital health education?	Infrastructure (human resources)	[96]
What are the workforce resources needed for health professions' digital education?	Infrastructure (human resources)	[61]

Research questions identified from included systematic reviews	Conceptual framework domain (subdomain)	Systematic reviews' references
What type of instructional design is used in the effective digital education of health professions education?	Education (modality)	[47,57,83,87,94]
Which components of digital health education (eg, interactivity and feedback) contribute to enhanced learning outcomes?	Education (modality)	[45,52,58,67,97]
What is the optimal use of video-assisted debriefing for health professionals' simulation-based training?	Education (modality)	[98]
How does the design of digital education interventions (eg, format and modality used) in health professions education and training curriculum affect learning outcomes?	Education (modality)	[34,42,53,64,74,78,93]
Can digital simulation-based training be used to train nontechnical skills in health professionals?	Education (modality)	[44,69]
What is the effectiveness of digital education (mixed or single modality) compared with nondigital education to deliver health professions education?	Education (modality)	[42,71,98]
Can digital education complement (ie, blended) or substitute traditional education for health professionals?	Education (modality)	[54,69,99,100]
Does digital simulation-based psychomotor skills training provide any benefit to the medical trainee?	Education (modality)	[46]
What are the barriers to obtaining digital education materials for health professions education training, and how can they be overcome?	Education (content)	[66]
What content should be included in debriefing (eg, digital data) following simulation-based education to achieve improved clinical outcomes?	Education (content)	[47]
Can digital education be used to overcome challenges in delivering content-specific topics for health professions education (eg, surgical training in rare pathologic states)?	Education (content)	[84,89]
Can digital education be designed to achieve learning outcomes denoted in the Kirkpatrick model?	Education (instructional design)	[101]
What learning theories can be used to inform the development of effective digital health professions education?	Education (instructional design)	[13,14,55,59,63,82,93]
Is mastery learning via digital education more or as effective as traditional education in terms of clinical psychomotor skills improvement?	Education (instructional design)	[41,47,48,53,58-60,78,102]
Is spacing digital simulation-based training more or as effective as traditional education in clinical psychomotor skills development?	Education (instructional design)	[41,47,48,53,58-60,78,102]
How does the frequency and duration of digital simulation-based psychomotor skills training affect health professionals' skills transfer to the clinical setting?	Education (instructional design)	[41,47,48,53,58-60,78,102]
What are the optimal duration, frequency, and intensity of digital health professions education programs to affect the learning and clinical outcomes of health professionals?	Education (instructional design)	[43,54,62,72,83,94,103]
What pedagogy should be used in the digital education of health professionals to improve their knowledge and skills?	Education (instructional design)	[11,14,42,95,104]
What is the effectiveness of using digital education to train and assess nontechnical skills in health care professionals?	Education (instructional design)	[71,87]
What is the effectiveness of digital problem-based learning in health professions education?	Education (instructional design)	[34]
How does the interactivity of digital education programs affect the learning and clinical outcomes of health professionals?	Education (engagement)	[53,62,80,91]
What is the minimal level of haptic feedback required in digital simulation-based training programs to improve health professionals' psychomotor skills?	Education (engagement)	[64]
What are learners' acceptability of digital education with different levels of interactivity?	Education (engagement)	[77]
Which performance metrics or measurement instrument should be used to assess health professionals' knowledge, skills, attitudes, satisfaction, and clinical outcomes from digital technology-based training programs?	Education (assessment)	[12, 14, 44, 45, 51-53, 60, 62, 64, 67-71, 73-75, 77, 78, 83, 87, 90, 92, 93, 95, 102, 103, 105]

Research questions identified from included systematic reviews	Conceptual framework domain (subdomain)	Systematic reviews' references
What is the ideal approach to assessing health professionals' knowledge, skills, attitudes, satisfaction, and clinical outcomes from digital technology-based education and training programs?	Education (assessment)	[12, 14, 44, 45, 51-53, 60, 62, 64, 67-71, 73-75, 77, 78, 83, 87, 90, 92, 93, 95, 102, 103, 105]
Should the evaluation of digital health education include behavior and clinical outcomes?	Education (assessment)	[12, 14, 44, 45, 51-53, 60, 62, 64, 67-71, 73-75, 77, 78, 83, 87, 90, 92, 93, 95, 102, 103, 105]
What is the impact of digital simulation-based training on clinical outcomes in the short and long term?	Education (assessment)	[71,106]
How should learning outcomes in the field of digital health professions education be defined and standardized?	Education (assessment)	[92]
How does the use of digital education affect health professionals' clinical decision-making at the point of care?	Education (assessment)	[68]
How do health professionals' prior learning experiences influence the topics that will benefit from the use of digital education?	Learner	[107]
What are health professionals' attitudes toward digital delivery of education and training programs?	Learner	[16,65,85,91]
What are health care professionals' learning needs, and can they be met by the use of digital simulation training?	Learner	[44]
What are the methodological requirements for studies assessing digital health education?	Research	[12, 16, 48, 53, 58, 59, 63, 65, 66, 70, 71, 74, 82, 83, 92, 95, 97]
How should studies on digital health professions education be reported?	Research	[12, 16, 48, 53, 58, 59, 63, 65, 66, 70, 71, 74, 82, 83, 92, 95, 97]
How should studies of digital health professions education be designed to ensure the generalizability of their findings across different settings?	Research	[12, 16, 48, 53, 58, 59, 63, 65, 66, 70, 71, 74, 82, 83, 92, 95, 97]
What are the barriers and facilitators that affect the continued adoption of digital tools in health professions education?	Context, education, infrastructure, and learner	[68]

Discussion

Principal Findings

We present an evidence map of 77 systematic reviews on digital education for health professionals published between 2014 and July 2020. The reviews mostly focused on the effectiveness of various digital education modalities in surgery, health professions education in general, and nursing. Most reviews have focused on online and offline learning. Only a few reviews focused on other digital education modalities such as m-Learning, VR, digital game-based learning, and virtual patients. We developed a novel conceptual framework outlining key components of digital health professions education, namely context, infrastructure, education, learner, research, and quality assurance. Within these reviews, we identified 61 unique recommendations (questions) for future research, focusing primarily on digital education modality, instructional design, and assessment.

Limitations and Strengths

Our study has some limitations. First, to cover the most recent evidence in the field of digital education for health care professionals, we excluded studies before 2014; earlier reviews might have identified important research questions that remain unanswered. Our focus on systematic reviews also excluded other article types, such as editorials or viewpoints, which might have identified additional research questions. Second, although

extraction and classification of research questions were done in duplicate and using a standardized approach, other classifications could be justified in some instances, which implies a degree of imprecision in the reported frequencies of specific questions. Moreover, our method did not allow us to prioritize the numerous research questions; such prioritization would require input from a representative group of experts and could be the focus of a future study. Third, there are overlaps among different concepts specified within this conceptual framework, which could be delineated and presented differently depending on potential chosen emphasis or entry points. Fourth, reviews classified as *online education* varied substantially in their inclusion of other modalities (eg, some expressly excluded modalities such as virtual patients, digital games, or massive open online courses, whereas others included these and other modalities). Finally, our novel conceptual framework may require revision as our understanding of this field matures and evolves, additional evidence accumulates, and new technologies emerge.

Our study also has several strengths, such as a thorough literature search for relevant studies, encompassing several indexed and gray literature databases without restrictions. We followed an established evidence map methodology and performed the steps in duplicate and independently [17]. In the development of our conceptual framework, we drew from the existing frameworks, our previous work, and discussions with experts.

Integration With Prior Work

We did not find other frameworks presenting a high-level overview of the use and implementation of digital education for health professionals. Therefore, we drew from the general digital education literature and found several relevant frameworks. (Multimedia Appendix 5). The included frameworks, while providing an overview of digital education, often had additional objectives such as exploring the role of specific stakeholders (eg, the private sector or the ministries), identifying barriers to adoption, or analyzing a particular digital education aspect, setting, or configuration [33,35-37,113-117]. Our framework complements other frameworks by pulling together domains previously presented only in isolation and by adding novel subdomains such as the impact of training levels, the role of regulations and accreditations, and the importance of physical infrastructure (Multimedia Appendix 4 [11-16,32,34,40-95,98-107]).

Several viewpoint articles have offered research agendas for digital health professions education [28,118-120]. They focus primarily on the design of interventions and research studies in this field, which correspond to the domains of education (modality, instruction design, assessment, and engagement), research (quality of methods and reporting), and context (setting and level of education) in our framework. The agendas espoused in these viewpoints include questions that probe more narrowly and deeply on specific issues relevant to the design and focus of future studies (eg, the choice of comparison intervention and avoidance of confounding, integration of digital education across different institutions, and the need for interdisciplinary collaboration). Our framework was intentionally broad and comprehensive and enabled us to accommodate a variety of additional questions on previously neglected topics.

Implications for Research and Practice

Most reviews in our evidence map focused on the effectiveness of digital education interventions and rarely addressed issues around their implementation and adoption. These reviews also mostly compared the effectiveness of digital interventions with that of nondigital education. Findings from studies comparing digital and nondigital education have limited generalizability as these studies cannot account for variance within and between these 2 educational formats [118]. Future research should

compare different digital education modalities as such studies are more likely to generate meaningful, generalizable findings. It should also aim to explore potential challenges related to the implementation and adoption of digital education interventions in different settings.

There is also a need for more methodologically robust research and clearer terminology in this field. The quality of the evidence, as reported in the included reviews, was relatively low, with a limited number of studies measuring skills and knowledge retention. Furthermore, it was at times difficult to determine which modality (or modalities) the included reviews focused on because of poorly explained inclusion criteria. Such ambiguity was particularly common in reviews on e-learning and blended, online, and internet-based education.

We also express concerns about the paucity of studies from low- and middle-income countries. Such countries could greatly benefit from digital education, especially by using free or low-cost education (eg, massive open online courses). Although some research findings have a universal application (eg, fundamental principles of effective learning), others (such as implementation, infrastructure, and learners) are more context specific. Given the presence of unique needs of low- and middle-income countries (eg, distinct content priorities, learner demographics, and infrastructure), we urge more research in these contexts.

Our conceptual framework will benefit researchers, funding agencies, and educators, among others. The specific questions identified and classified according to this framework provide a map for future research and can help prioritize original research studies and guide the planning of new or updated systematic reviews. We encourage investigators to broadly consider the questions we identified in this evidence map, especially those specific to areas previously less studied, such as infrastructure, learners, or quality assurance in digital education. Our framework can also be used by funding agencies to better understand the limitations of the existing research and identify areas with limited evidence with the aim of informing their funding calls in this field. Finally, this framework can encourage those developing new courses to anticipate and plan for issues that are important but might be inadvertently overlooked, such as the digital education context, infrastructure, and learners.

Acknowledgments

The authors would like to acknowledge the funding support from Nanyang Technological University, Singapore. The authors would also like to acknowledge Ms Thai My Linh and Ms Christina Tan En Hui for their help with the validation of the findings, Ms Chong Zhuo En for help with data compilation, and Ms Yasmin Lynda Munro for providing advice on our search strategy. Although this review includes the findings from the systematic reviews published by members of the Digital Health Education Collaboration, all reviews were given equal treatment during the review process conducted for this publication. The authors alone are responsible for the views expressed in this paper and do not necessarily represent the views, decisions, or policies of the institutions with which they are affiliated.

Authors' Contributions

JC, LTC, and OA conceived the idea for the review. SP and BMK screened the studies, extracted the data, and synthesized the findings. LTC, SP, and BMK wrote the manuscript. JC, DAC, VW, RA, AM, JJ, RMJJvdK, MM, FVW, ML, NC, OA, and CGP provided insightful feedback on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

MEDLINE (Ovid) search strategy.

[[DOCX File , 18 KB - jmir_v24i3e31977_app1.docx](#)]

Multimedia Appendix 2

Educational outcomes reported in the included systematic reviews and their definitions.

[[DOCX File , 14 KB - jmir_v24i3e31977_app2.docx](#)]

Multimedia Appendix 3

Characteristics of the included systematic reviews.

[[DOCX File , 92 KB - jmir_v24i3e31977_app3.docx](#)]

Multimedia Appendix 4

Research questions identified in the included systematic reviews.

[[DOCX File , 193 KB - jmir_v24i3e31977_app4.docx](#)]

Multimedia Appendix 5

Overview of conceptual frameworks on the implementation and adoption of digital education.

[[DOCX File , 17 KB - jmir_v24i3e31977_app5.docx](#)]

References

1. Working for health and growth: investing in the health workforce. World Health Organization. 2016. URL: <https://www.who.int/publications/i/item/9789241511308> [accessed 2022-02-19]
2. Health workforce. World Health Organization. 2019. URL: https://www.who.int/health-topics/health-workforce#tab=tab_1 [accessed 2022-02-19]
3. The world health report 2006. Working together for health. World Health Organization. 2006. URL: https://www.who.int/workforcealliance/knowledge/resources/whreport_2006/en/ [accessed 2022-02-19]
4. Scheffler R, Cometto G, Tulenko K, Bruckner T, Liu J, Keuffel EL, et al. Health workforce requirements for universal health coverage and the Sustainable Development Goals – background paper No.1 to the WHO Global Strategy on Human Resources for Health: workforce 2030. Human Resources for Health Observer Series No 17. Geneva: World Health Organization; 2016. URL: <https://apps.who.int/iris/bitstream/handle/10665/250330/9789241511407-eng.pdf> [accessed 2022-02-19]
5. Liu JX, Goryakin Y, Maeda A, Bruckner T, Scheffler R. Global health workforce labor market projections for 2030. Hum Resour Health 2017;15(1):11 [FREE Full text] [doi: [10.1186/s12960-017-0187-2](https://doi.org/10.1186/s12960-017-0187-2)] [Medline: [28159017](https://pubmed.ncbi.nlm.nih.gov/28159017/)]
6. Tumlinson K, Jaff D, Stilwell B, Onyango DO, Leonard KL. Reforming medical education admission and training in low- and middle-income countries: who gets admitted and why it matters. Hum Resour Health 2019;17(1):91 [FREE Full text] [doi: [10.1186/s12960-019-0426-9](https://doi.org/10.1186/s12960-019-0426-9)] [Medline: [31791358](https://pubmed.ncbi.nlm.nih.gov/31791358/)]
7. Hippe DS, Umoren RA, McGee A, Bucher SL, Bresnahan BW. A targeted systematic review of cost analyses for implementation of simulation-based education in healthcare. SAGE Open Med 2020;8:2050312120913451 [FREE Full text] [doi: [10.1177/2050312120913451](https://doi.org/10.1177/2050312120913451)] [Medline: [32231781](https://pubmed.ncbi.nlm.nih.gov/32231781/)]
8. Guze PA. Using technology to meet the challenges of medical education. Trans Am Clin Climatol Assoc 2015;126:260-270 [FREE Full text] [Medline: [26330687](https://pubmed.ncbi.nlm.nih.gov/26330687/)]
9. Goh PS, Sandars J. A vision of the use of technology in medical education after the COVID-19 pandemic. MedEdPublish 2020;9(1):49. [doi: [10.15694/mep.2020.000049.1](https://doi.org/10.15694/mep.2020.000049.1)]
10. Arney LE, Hosman L. The centrality of electricity to ICT use in low-income countries. Telecomm Policy 2016;40(7):617-627. [doi: [10.1016/j.telpol.2015.08.005](https://doi.org/10.1016/j.telpol.2015.08.005)]
11. Dunleavy G, Nikolaou CK, Nifakos S, Atun R, Law GC, Tudor Car L. Mobile digital education for health professions: systematic review and meta-analysis by the digital health education collaboration. J Med Internet Res 2019;21(2):e12937 [FREE Full text] [doi: [10.2196/12937](https://doi.org/10.2196/12937)] [Medline: [30747711](https://pubmed.ncbi.nlm.nih.gov/30747711/)]
12. George PP, Papachristou N, Belisario JM, Wang W, Wark PA, Cotic Z, et al. Online eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. J Glob Health 2014;4(1):010406 [FREE Full text] [doi: [10.7189/jogh.04.010406](https://doi.org/10.7189/jogh.04.010406)] [Medline: [24976965](https://pubmed.ncbi.nlm.nih.gov/24976965/)]

13. George PP, Zhabenko O, Kyaw BM, Antoniou P, Posadzki P, Saxena N, et al. Online digital education for postregistration training of medical doctors: systematic review by the digital health education collaboration. *J Med Internet Res* 2019;21(2):e13269 [FREE Full text] [doi: [10.2196/13269](https://doi.org/10.2196/13269)] [Medline: [30801252](https://pubmed.ncbi.nlm.nih.gov/30801252/)]
14. Gentry SV, Gauthier A, L'Estrade Ehrstrom B, Wortley D, Lilienthal A, Tudor Car L, et al. Serious gaming and gamification education in health professions: systematic review. *J Med Internet Res* 2019;21(3):e12994 [FREE Full text] [doi: [10.2196/12994](https://doi.org/10.2196/12994)] [Medline: [30920375](https://pubmed.ncbi.nlm.nih.gov/30920375/)]
15. Posadzki P, Bala MM, Kyaw BM, Semwal M, Divakar U, Koperny M, et al. Offline digital education for postregistration health professions: systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res* 2019;21(4):e12968 [FREE Full text] [doi: [10.2196/12968](https://doi.org/10.2196/12968)] [Medline: [31017584](https://pubmed.ncbi.nlm.nih.gov/31017584/)]
16. Kyaw BM, Saxena N, Posadzki P, Vseteckova J, Nikolaou CK, George PP, et al. Virtual reality for health professions education: systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res* 2019;21(1):e12959 [FREE Full text] [doi: [10.2196/12959](https://doi.org/10.2196/12959)] [Medline: [30668519](https://pubmed.ncbi.nlm.nih.gov/30668519/)]
17. Singh K, Ansari M, Galipeau J, Garrity C, Keely E, Malcolm J, et al. An evidence map of systematic reviews to inform interventions in prediabetes. *Can J Diabetes* 2012;36(5):281-291. [doi: [10.1016/j.cjcd.2012.06.004](https://doi.org/10.1016/j.cjcd.2012.06.004)]
18. Jabareen Y. Building a conceptual framework: philosophy, definitions, and procedure. *Int J Qual Methods* 2009;8(4):49-62. [doi: [10.1177/160940690900800406](https://doi.org/10.1177/160940690900800406)]
19. ISCED fields of education and training 2013 (ISCED-F 2013). UNESCO Institute for Statistics. Montreal: UNESCO Institute for Statistics; 2014. URL: <http://uis.unesco.org/sites/default/files/documents/isced-fields-of-education-and-training-2013-en.pdf> [accessed 2022-02-21]
20. 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;21(2):e12913 [FREE Full text] [doi: [10.2196/12913](https://doi.org/10.2196/12913)] [Medline: [30762583](https://pubmed.ncbi.nlm.nih.gov/30762583/)]
21. Hervatis V, Kyaw BM, Semwal M, Dunleavy G, Tudor Car L, Zary N, et al. Offline and computer-based eLearning interventions for medical students' education. *Cochrane Database Syst Rev* 2018;2018(10):CD012149. [doi: [10.1002/14651858.CD012149.pub2](https://doi.org/10.1002/14651858.CD012149.pub2)]
22. Paul P, Toon E, Hadadgar A, Jirwe M, Saxena N, Lim KT, et al. Online- and local area network (LAN)-based eLearning interventions for medical doctors' education. *Cochrane Database Syst Rev* 2018;2018(10):CD012108. [doi: [10.1002/14651858.cd012108.pub2](https://doi.org/10.1002/14651858.cd012108.pub2)]
23. Cook DA, Levinson AJ, Garside S, Dupras DM, Erwin PJ, Montori VM. Instructional design variations in internet-based learning for health professions education: a systematic review and meta-analysis. *Acad Med* 2010;85(5):909-922. [doi: [10.1097/ACM.0b013e3181d6c319](https://doi.org/10.1097/ACM.0b013e3181d6c319)] [Medline: [20520049](https://pubmed.ncbi.nlm.nih.gov/20520049/)]
24. Stuchlikova L, Kósa A. Massive open online courses - challenges and solutions in engineering education. In: *IEEE 11th International Conference on Emerging eLearning Technologies and Applications*. 2013 Presented at: ICETA '13; October 24-25, 2013; Stara Lesna, Slovakia p. 359-364. [doi: [10.1109/iceta.2013.6674459](https://doi.org/10.1109/iceta.2013.6674459)]
25. Crompton H. A historical overview of M-learning: toward learner-centered education. In: Berge ZL, Muilenburg LY, editors. *Handbook of mobile learning*. Milton Park, UK: Routledge; 2013.
26. Gentry S, L'Estrade Ehrstrom B, Gauthier A, Alvarez J, Wortley D, van Rijswijk J, et al. Serious gaming and gamification interventions for health professional education. *Cochrane Database Syst Rev* 2018;2018(6):CD012209. [doi: [10.1002/14651858.CD012209.pub2](https://doi.org/10.1002/14651858.CD012209.pub2)]
27. Saxena N, Kyaw BM, Vseteckova J, Dev P, Paul P, Lim KT, et al. Virtual reality environments for health professional education. *Cochrane Database Syst Rev* 2016;2016(2):CD012090. [doi: [10.1002/14651858.cd012090](https://doi.org/10.1002/14651858.cd012090)]
28. Cook DA, Triola MM. Virtual patients: a critical literature review and proposed next steps. *Med Educ* 2009;43(4):303-311. [doi: [10.1111/j.1365-2923.2008.03286.x](https://doi.org/10.1111/j.1365-2923.2008.03286.x)] [Medline: [19335571](https://pubmed.ncbi.nlm.nih.gov/19335571/)]
29. Lopreiato JO. *Healthcare simulation dictionary*. Rockville, MD: Agency for Healthcare and Quality; 2016.
30. Tiven MB, Fuchs ER, Bazari A, Macquarrie A. Evaluating global digital education: student outcomes framework. Organisation for Economic co-operation and development. 2018. URL: <https://www.oecd.org/pisa/Evaluating-Global-Digital-Education-Student-Outcomes-Framework.pdf> [accessed 2022-02-21]
31. Digital learning and teaching framework. RMIT University. 2017. URL: <https://cpb-ap-se2.wpmucdn.com/sites.rmit.edu.au/dist/c/16/files/2017/03/Draft-Digital-Learning-Framework-28.3.2017-vh7abk.pdf> [accessed 2022-02-21]
32. Kyaw BM, Tudor Car L, van Galen LS, van Agtmael MA, Costelloe CE, Ajuebor O, et al. Health professions digital education on antibiotic management: systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res* 2019;21(9):e14984 [FREE Full text] [doi: [10.2196/14984](https://doi.org/10.2196/14984)] [Medline: [31516125](https://pubmed.ncbi.nlm.nih.gov/31516125/)]
33. Tudor Car L, Kyaw BM, Atun R. The role of eLearning in health management and leadership capacity building in health system: a systematic review. *Hum Resour Health* 2018;16(1):44 [FREE Full text] [doi: [10.1186/s12960-018-0305-9](https://doi.org/10.1186/s12960-018-0305-9)] [Medline: [30176899](https://pubmed.ncbi.nlm.nih.gov/30176899/)]
34. 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;21(2):e12945 [FREE Full text] [doi: [10.2196/12945](https://doi.org/10.2196/12945)] [Medline: [30816846](https://pubmed.ncbi.nlm.nih.gov/30816846/)]

35. Khan BH. The global e-Learning framework. In: Mishra S, editor. E-Learning (STRIDE Handbook 8). New Delhi, India: Staff Training and Research Institute of Distance Education (STRIDE), IGNOU; 2009:42-51.
36. Aparicio M, Bacao F, Oliveira T. An e-Learning theoretical framework. *J Educ Tech Syst* 2016;19(1):292-307.
37. Andersson A, Grönlund Å. A conceptual framework for e-learning in developing countries: a critical review of research challenges. *Electron J Inf Syst Dev Ctries* 2009;38(1):1-16 [FREE Full text] [doi: [10.1002/j.1681-4835.2009.tb00271.x](https://doi.org/10.1002/j.1681-4835.2009.tb00271.x)]
38. Wisdom JP, Chor KH, Hoagwood KE, Horwitz SM. Innovation adoption: a review of theories and constructs. *Adm Policy Ment Health* 2014;41(4):480-502 [FREE Full text] [doi: [10.1007/s10488-013-0486-4](https://doi.org/10.1007/s10488-013-0486-4)] [Medline: [23549911](https://pubmed.ncbi.nlm.nih.gov/23549911/)]
39. 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;19(11):e367 [FREE Full text] [doi: [10.2196/jmir.8775](https://doi.org/10.2196/jmir.8775)] [Medline: [29092808](https://pubmed.ncbi.nlm.nih.gov/29092808/)]
40. Kyaw BM, Posadzki P, Dunleavy G, Semwal M, Divakar U, Hervatis V, et al. Offline digital education for medical students: systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res* 2019;21(3):e13165 [FREE Full text] [doi: [10.2196/13165](https://doi.org/10.2196/13165)] [Medline: [30907731](https://pubmed.ncbi.nlm.nih.gov/30907731/)]
41. Jayakumar N, Brunckhorst O, Dasgupta P, Khan MS, Ahmed K. e-Learning in surgical education: a systematic review. *J Surg Educ* 2015;72(6):1145-1157. [doi: [10.1016/j.jsurg.2015.05.008](https://doi.org/10.1016/j.jsurg.2015.05.008)] [Medline: [26111822](https://pubmed.ncbi.nlm.nih.gov/26111822/)]
42. Pei L, Wu H. Does online learning work better than offline learning in undergraduate medical education? A systematic review and meta-analysis. *Med Educ Online* 2019;24(1):1666538 [FREE Full text] [doi: [10.1080/10872981.2019.1666538](https://doi.org/10.1080/10872981.2019.1666538)] [Medline: [31526248](https://pubmed.ncbi.nlm.nih.gov/31526248/)]
43. Viljoen CA, Scott Millar R, Engel ME, Shelton M, Burch V. Is computer-assisted instruction more effective than other educational methods in achieving ECG competence amongst medical students and residents? A systematic review and meta-analysis. *BMJ Open* 2019;9(11):e028800 [FREE Full text] [doi: [10.1136/bmjopen-2018-028800](https://doi.org/10.1136/bmjopen-2018-028800)] [Medline: [31740464](https://pubmed.ncbi.nlm.nih.gov/31740464/)]
44. Boutefnouchet T, Laios T. Transfer of arthroscopic skills from computer simulation training to the operating theatre: a review of evidence from two randomised controlled studies. *SICOT J* 2016;2:4 [FREE Full text] [doi: [10.1051/sicotj/2015039](https://doi.org/10.1051/sicotj/2015039)] [Medline: [27163093](https://pubmed.ncbi.nlm.nih.gov/27163093/)]
45. Lapiere A, Bouferguene S, Gauvin-Lepage J, Lavoie P, Arbour C. Effectiveness of interprofessional manikin-based simulation training on teamwork among real teams during trauma resuscitation in adult emergency departments: a systematic review. *Simul Healthc* 2020;15(6):409-421. [doi: [10.1097/SIH.0000000000000443](https://doi.org/10.1097/SIH.0000000000000443)] [Medline: [32218090](https://pubmed.ncbi.nlm.nih.gov/32218090/)]
46. Dromey BP, Peebles DM, Stoyanov DV. A systematic review and meta-analysis of the use of high-fidelity simulation in obstetric ultrasound. *Simul Healthc* 2021;16(1):52-59 [FREE Full text] [doi: [10.1097/SIH.0000000000000485](https://doi.org/10.1097/SIH.0000000000000485)] [Medline: [32675735](https://pubmed.ncbi.nlm.nih.gov/32675735/)]
47. Cheng A, Lockey A, Bhanji F, Lin Y, Hunt EA, Lang E. The use of high-fidelity manikins for advanced life support training--a systematic review and meta-analysis. *Resuscitation* 2015;93:142-149. [doi: [10.1016/j.resuscitation.2015.04.004](https://doi.org/10.1016/j.resuscitation.2015.04.004)] [Medline: [25888241](https://pubmed.ncbi.nlm.nih.gov/25888241/)]
48. Mok HT, So CF, Chung JW. Effectiveness of high-fidelity patient simulation in teaching clinical reasoning skills. *Clin Simul Nurs* 2016;12(10):453-467. [doi: [10.1016/j.ecns.2016.06.003](https://doi.org/10.1016/j.ecns.2016.06.003)]
49. Santos GN, Leite AF, Figueiredo PT, Pimentel NM, Flores-Mir C, de Melo NS, et al. Effectiveness of e-learning in oral radiology education: a systematic review. *J Dent Educ* 2016;80(9):1126-1139. [Medline: [27587580](https://pubmed.ncbi.nlm.nih.gov/27587580/)]
50. Tomesko J, Touger-Decker R, Dreker M, Zelig R, Parrott JS. The effectiveness of computer-assisted instruction to teach physical examination to students and trainees in the health sciences professions: a systematic review and meta-analysis. *J Med Educ Curric Dev* 2017;4:2382120517720428 [FREE Full text] [doi: [10.1177/2382120517720428](https://doi.org/10.1177/2382120517720428)] [Medline: [29349338](https://pubmed.ncbi.nlm.nih.gov/29349338/)]
51. Ahmet A, Gamze K, Rustem M, Sezen KA. Is video-based education an effective method in surgical education? A systematic review. *J Surg Educ* 2018;75(5):1150-1158. [doi: [10.1016/j.jsurg.2018.01.014](https://doi.org/10.1016/j.jsurg.2018.01.014)] [Medline: [29449162](https://pubmed.ncbi.nlm.nih.gov/29449162/)]
52. Lioffi C, Failo A, Schoth DE, Williams G, Howard RF. The effectiveness of online pain resources for health professionals: a systematic review with subset meta-analysis of educational intervention studies. *Pain* 2018;159(4):631-643. [doi: [10.1097/j.pain.0000000000001146](https://doi.org/10.1097/j.pain.0000000000001146)] [Medline: [29319610](https://pubmed.ncbi.nlm.nih.gov/29319610/)]
53. Tudor Car L, Soong A, Kyaw BM, Chua KL, Low-Beer N, Majeed A. Health professions digital education on clinical practice guidelines: a systematic review by Digital Health Education collaboration. *BMC Med* 2019;17(1):139 [FREE Full text] [doi: [10.1186/s12916-019-1370-1](https://doi.org/10.1186/s12916-019-1370-1)] [Medline: [31315642](https://pubmed.ncbi.nlm.nih.gov/31315642/)]
54. Foronda CL, Fernandez-Burgos M, Nadeau C, Kelley CN, Henry MN. Virtual simulation in nursing education: a systematic review spanning 1996 to 2018. *Simul Healthc* 2020;15(1):46-54. [doi: [10.1097/SIH.0000000000000411](https://doi.org/10.1097/SIH.0000000000000411)] [Medline: [32028447](https://pubmed.ncbi.nlm.nih.gov/32028447/)]
55. Martinengo L, Yeo NJ, Markandran KD, Olsson M, Kyaw BM, Car LT. Digital health professions education on chronic wound management: a systematic review. *Int J Nurs Stud* 2020;104:103512. [doi: [10.1016/j.ijnurstu.2019.103512](https://doi.org/10.1016/j.ijnurstu.2019.103512)] [Medline: [32086027](https://pubmed.ncbi.nlm.nih.gov/32086027/)]
56. Brusamento S, Kyaw BM, Whiting P, Li L, Tudor Car L. Digital health professions education in the field of pediatrics: systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res* 2019;21(9):e14231 [FREE Full text] [doi: [10.2196/14231](https://doi.org/10.2196/14231)] [Medline: [31573906](https://pubmed.ncbi.nlm.nih.gov/31573906/)]
57. Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. The effectiveness of blended learning in health professions: systematic review and meta-analysis. *J Med Internet Res* 2016;18(1):e2 [FREE Full text] [doi: [10.2196/jmir.4807](https://doi.org/10.2196/jmir.4807)] [Medline: [26729058](https://pubmed.ncbi.nlm.nih.gov/26729058/)]

58. Alaker M, Wynn GR, Arulampalam T. *Int J Surg* 2016;29:85-94 [FREE Full text] [doi: [10.1016/j.ijso.2016.03.034](https://doi.org/10.1016/j.ijso.2016.03.034)] [Medline: [26992652](https://pubmed.ncbi.nlm.nih.gov/26992652/)]
59. Khan R, Plahouras J, Johnston BC, Scaffidi MA, Grover SC, Walsh CM. Virtual reality simulation training for health professions trainees in gastrointestinal endoscopy. *Cochrane Database Syst Rev* 2018;8:CD008237 [FREE Full text] [doi: [10.1002/14651858.CD008237.pub3](https://doi.org/10.1002/14651858.CD008237.pub3)] [Medline: [30117156](https://pubmed.ncbi.nlm.nih.gov/30117156/)]
60. Koohestani HR, Soltani Arabshahi SK, Fata L, Ahmadi F. The educational effects of mobile learning on students of medical sciences: a systematic review in experimental studies. *J Adv Med Educ Prof* 2018;6(2):58-69 [FREE Full text] [Medline: [29607333](https://pubmed.ncbi.nlm.nih.gov/29607333/)]
61. Rambarat CA, Merritt JM, Norton HF, Black E, Winchester DE. Using Simulation to Teach Echocardiography: a Systematic Review. *Simul Healthc* 2018;13(6):413-419. [doi: [10.1097/SIH.0000000000000351](https://doi.org/10.1097/SIH.0000000000000351)] [Medline: [30520805](https://pubmed.ncbi.nlm.nih.gov/30520805/)]
62. Vaona A, Banzi R, Kwag KH, Rigon G, Cereda D, Pecoraro V, et al. E-learning for health professionals. *Cochrane Database Syst Rev* 2018;1(1):CD011736 [FREE Full text] [doi: [10.1002/14651858.CD011736.pub2](https://doi.org/10.1002/14651858.CD011736.pub2)] [Medline: [29355907](https://pubmed.ncbi.nlm.nih.gov/29355907/)]
63. Xu X, Posadzki PP, Lee GE, Car J, Smith HE. Digital education for health professions in the field of dermatology: a systematic review by digital health education collaboration. *Acta Derm Venereol* 2019;99(2):133-138 [FREE Full text] [doi: [10.2340/00015555-3068](https://doi.org/10.2340/00015555-3068)] [Medline: [30320871](https://pubmed.ncbi.nlm.nih.gov/30320871/)]
64. Rangarajan K, Davis H, Pucher PH. Systematic review of virtual haptics in surgical simulation: a valid educational tool? *J Surg Educ* 2020;77(2):337-347. [doi: [10.1016/j.jsurg.2019.09.006](https://doi.org/10.1016/j.jsurg.2019.09.006)] [Medline: [31564519](https://pubmed.ncbi.nlm.nih.gov/31564519/)]
65. Rourke S. How does virtual reality simulation compare to simulated practice in the acquisition of clinical psychomotor skills for pre-registration student nurses? A systematic review. *Int J Nurs Stud* 2020;102:103466. [doi: [10.1016/j.ijnurstu.2019.103466](https://doi.org/10.1016/j.ijnurstu.2019.103466)] [Medline: [31783192](https://pubmed.ncbi.nlm.nih.gov/31783192/)]
66. Rasmussen K, Belisario JM, Wark PA, Molina JA, Loong SL, Cotic Z, et al. Offline eLearning for undergraduates in health professions: a systematic review of the impact on knowledge, skills, attitudes and satisfaction. *J Glob Health* 2014;4(1):010405 [FREE Full text] [doi: [10.7189/jogh.04.010405](https://doi.org/10.7189/jogh.04.010405)] [Medline: [24976964](https://pubmed.ncbi.nlm.nih.gov/24976964/)]
67. Thepwongsa I, Kirby CN, Schattner P, Piterman L. Online continuing medical education (CME) for GPs: does it work? A systematic review. *Aust Fam Physician* 2014;43(10):717-721 [FREE Full text] [Medline: [25286431](https://pubmed.ncbi.nlm.nih.gov/25286431/)]
68. Mickan S, Atherton H, Roberts NW, Heneghan C, Tilson JK. Use of handheld computers in clinical practice: a systematic review. *BMC Med Inform Decis Mak* 2014;14:56 [FREE Full text] [doi: [10.1186/1472-6947-14-56](https://doi.org/10.1186/1472-6947-14-56)] [Medline: [24998515](https://pubmed.ncbi.nlm.nih.gov/24998515/)]
69. Pirochchai P, Avery A, Laopaiboon M, Kennedy G, O'Leary S. Virtual reality training for improving the skills needed for performing surgery of the ear, nose or throat. *Cochrane Database Syst Rev* 2015(9):CD010198. [doi: [10.1002/14651858.CD010198.pub2](https://doi.org/10.1002/14651858.CD010198.pub2)] [Medline: [26352008](https://pubmed.ncbi.nlm.nih.gov/26352008/)]
70. Aim F, Lonjon G, Hannouche D, Nizard R. Effectiveness of virtual reality training in orthopaedic surgery. *Arthroscopy* 2016;32(1):224-232. [doi: [10.1016/j.arthro.2015.07.023](https://doi.org/10.1016/j.arthro.2015.07.023)] [Medline: [26412672](https://pubmed.ncbi.nlm.nih.gov/26412672/)]
71. Pfandler M, Lazarovici M, Stefan P, Wucherer P, Weigl M. Virtual reality-based simulators for spine surgery: a systematic review. *Spine J* 2017;17(9):1352-1363. [doi: [10.1016/j.spinee.2017.05.016](https://doi.org/10.1016/j.spinee.2017.05.016)] [Medline: [28571789](https://pubmed.ncbi.nlm.nih.gov/28571789/)]
72. Kang J, Seomun G. Evaluating web-based nursing education's effects: a systematic review and meta-analysis. *West J Nurs Res* 2018;40(11):1677-1697. [doi: [10.1177/0193945917729160](https://doi.org/10.1177/0193945917729160)] [Medline: [28875770](https://pubmed.ncbi.nlm.nih.gov/28875770/)]
73. Lee H, Min H, Oh SM, Shim K. Mobile technology in undergraduate nursing education: a systematic review. *Healthc Inform Res* 2018;24(2):97-108 [FREE Full text] [doi: [10.4258/hir.2018.24.2.97](https://doi.org/10.4258/hir.2018.24.2.97)] [Medline: [29770243](https://pubmed.ncbi.nlm.nih.gov/29770243/)]
74. Botelho MG, Agrawal KR, Bornstein MM. An systematic review of e-learning outcomes in undergraduate dental radiology curricula-levels of learning and implications for researchers and curriculum planners. *Dentomaxillofac Radiol* 2019;48(1):20180027 [FREE Full text] [doi: [10.1259/dmfr.20180027](https://doi.org/10.1259/dmfr.20180027)] [Medline: [30028185](https://pubmed.ncbi.nlm.nih.gov/30028185/)]
75. Kononowicz AA, Woodham LA, Edelbring S, Stathakarou N, Davies D, Saxena N, et al. Virtual patient simulations in health professions education: systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res* 2019;21(7):e14676 [FREE Full text] [doi: [10.2196/14676](https://doi.org/10.2196/14676)] [Medline: [31267981](https://pubmed.ncbi.nlm.nih.gov/31267981/)]
76. La Cerra C, Dante A, Caponnetto V, Franconi I, Gaxhja E, Petrucci C, et al. Effects of high-fidelity simulation based on life-threatening clinical condition scenarios on learning outcomes of undergraduate and postgraduate nursing students: a systematic review and meta-analysis. *BMJ Open* 2019;9(2):e025306 [FREE Full text] [doi: [10.1136/bmjopen-2018-025306](https://doi.org/10.1136/bmjopen-2018-025306)] [Medline: [30798316](https://pubmed.ncbi.nlm.nih.gov/30798316/)]
77. Nagendrababu V, Pulikkotil SJ, Sultan OS, Jayaraman J, Soh JA, Dummer PM. Effectiveness of technology-enhanced learning in Endodontic education: a systematic review and meta-analysis. *Int Endod J* 2019;52(2):181-192. [doi: [10.1111/iej.12995](https://doi.org/10.1111/iej.12995)] [Medline: [30099740](https://pubmed.ncbi.nlm.nih.gov/30099740/)]
78. Semwal M, Whiting P, Bajpai R, Bajpai S, Kyaw BM, Tudor Car L. Digital education for health professions on smoking cessation management: systematic review by the digital health education collaboration. *J Med Internet Res* 2019;21(3):e13000 [FREE Full text] [doi: [10.2196/13000](https://doi.org/10.2196/13000)] [Medline: [30829576](https://pubmed.ncbi.nlm.nih.gov/30829576/)]
79. Osborne B, Parange N, Thoires K. The effectiveness of the use of high fidelity simulators in obstetric ultrasound training: a systematic review. *Australas J Ultrasound Med* 2015;18(3):107-111 [FREE Full text] [doi: [10.1002/j.2205-0140.2015.tb00209.x](https://doi.org/10.1002/j.2205-0140.2015.tb00209.x)] [Medline: [28191251](https://pubmed.ncbi.nlm.nih.gov/28191251/)]

80. den Harder AM, Frijlingh M, Ravesloot CJ, Oosterbaan AE, van der Gijp A. The importance of human-computer interaction in radiology e-learning. *J Digit Imaging* 2016;29(2):195-205 [[FREE Full text](#)] [doi: [10.1007/s10278-015-9828-y](https://doi.org/10.1007/s10278-015-9828-y)] [Medline: [26464115](#)]
81. Glassman D, Yiasemidou M, Ishii H, Somani BK, Ahmed K, Biyani CS. Effect of playing video games on laparoscopic skills performance: a systematic review. *J Endourol* 2016;30(2):146-152. [doi: [10.1089/end.2015.0425](https://doi.org/10.1089/end.2015.0425)] [Medline: [26414043](#)]
82. Maertens H, Madani A, Landry T, Vermassen F, Van Herzeele I, Aggarwal R. Systematic review of e-learning for surgical training. *Br J Surg* 2016;103(11):1428-1437. [doi: [10.1002/bjs.10236](https://doi.org/10.1002/bjs.10236)] [Medline: [27537708](#)]
83. Sinclair PM, Kable A, Levett-Jones T, Booth D. The effectiveness of Internet-based e-learning on clinician behaviour and patient outcomes: a systematic review. *Int J Nurs Stud* 2016;57:70-81. [doi: [10.1016/j.ijnurstu.2016.01.011](https://doi.org/10.1016/j.ijnurstu.2016.01.011)] [Medline: [27045566](#)]
84. Lui JT, Hoy MY. Evaluating the effect of virtual reality temporal bone simulation on mastoidectomy performance: a meta-analysis. *Otolaryngol Head Neck Surg* 2017;156(6):1018-1024. [doi: [10.1177/0194599817698440](https://doi.org/10.1177/0194599817698440)] [Medline: [28418754](#)]
85. Rohwer A, Motaze NV, Rehfuess E, Young T. E-learning of evidence-based health care (EBHC) to increase EBHC competencies in healthcare professionals: a systematic review. *Campbell Syst Rev* 2017;13(1):1-147. [doi: [10.4073/csr.2017.4](https://doi.org/10.4073/csr.2017.4)]
86. Campbell K, Taylor V, Douglas S. Effectiveness of online cancer education for nurses and allied health professionals; a systematic review using Kirkpatrick evaluation framework. *J Cancer Educ* 2019;34(2):339-356. [doi: [10.1007/s13187-017-1308-2](https://doi.org/10.1007/s13187-017-1308-2)] [Medline: [29230687](#)]
87. Lee J, Kim H, Kim KH, Jung D, Jowsey T, Webster CS. Effective virtual patient simulators for medical communication training: a systematic review. *Med Educ* 2020;54(9):786-795. [doi: [10.1111/medu.14152](https://doi.org/10.1111/medu.14152)] [Medline: [32162355](#)]
88. Vitale SG, Caruso S, Vitagliano A, Vilos G, Di Gregorio LM, Zizolfi B, et al. The value of virtual reality simulators in hysteroscopy and training capacity: a systematic review. *Minim Invasive Ther Allied Technol* 2020;29(4):185-193. [doi: [10.1080/13645706.2019.1625404](https://doi.org/10.1080/13645706.2019.1625404)] [Medline: [31169414](#)]
89. Divakar U, Nazeha N, Posadzki P, Jarbrink K, Bajpai R, Ho AH, et al. Digital education of health professionals on the management of domestic violence: systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res* 2019;21(5):e13868 [[FREE Full text](#)] [doi: [10.2196/13868](https://doi.org/10.2196/13868)] [Medline: [31124462](#)]
90. Barteit S, Guzek D, Jahn A, Bärnighausen T, Jorge MM, Neuhann F. Evaluation of e-learning for medical education in low- and middle-income countries: a systematic review. *Comput Educ* 2020;145:103726 [[FREE Full text](#)] [doi: [10.1016/j.compedu.2019.103726](https://doi.org/10.1016/j.compedu.2019.103726)] [Medline: [32565611](#)]
91. Kyaw BM, Posadzki P, Paddock S, Car J, Campbell J, Tudor Car L. Effectiveness of digital education on communication skills among medical students: systematic review and meta-analysis by the digital health education collaboration. *J Med Internet Res* 2019;21(8):e12967 [[FREE Full text](#)] [doi: [10.2196/12967](https://doi.org/10.2196/12967)] [Medline: [31456579](#)]
92. Wahabi HA, Esmail SA, Bahkali KH, Titi MA, Amer YS, Fayed AA, et al. Medical doctors' offline computer-assisted digital education: systematic review by the digital health education collaboration. *J Med Internet Res* 2019;21(3):e12998 [[FREE Full text](#)] [doi: [10.2196/12998](https://doi.org/10.2196/12998)] [Medline: [30821689](#)]
93. Huang Z, Semwal M, Lee SY, Tee M, Ong W, Tan WS, et al. Digital health professions education on diabetes management: systematic review by the digital health education collaboration. *J Med Internet Res* 2019;21(2):e12997 [[FREE Full text](#)] [doi: [10.2196/12997](https://doi.org/10.2196/12997)] [Medline: [30789348](#)]
94. Rothschild P, Richardson A, Beltz J, Chakrabarti R. Effect of virtual reality simulation training on real-life cataract surgery complications: systematic literature review. *J Cataract Refract Surg* 2021;47(3):400-406. [doi: [10.1097/j.jcrs.000000000000323](https://doi.org/10.1097/j.jcrs.000000000000323)] [Medline: [32675654](#)]
95. Richmond H, Copey B, Hall AM, Davies D, Lamb SE. A systematic review and meta-analysis of online versus alternative methods for training licensed health care professionals to deliver clinical interventions. *BMC Med Educ* 2017;17(1):227 [[FREE Full text](#)] [doi: [10.1186/s12909-017-1047-4](https://doi.org/10.1186/s12909-017-1047-4)] [Medline: [29169393](#)]
96. Männistö M, Mikkonen K, Kuivila HM, Virtanen M, Kyngäs H, Kääriäinen M. Digital collaborative learning in nursing education: a systematic review. *Scand J Caring Sci* 2020;34(2):280-292. [doi: [10.1111/scs.12743](https://doi.org/10.1111/scs.12743)] [Medline: [31487063](#)]
97. Voutilainen A, Saaranen T, Sormunen M. Conventional vs. e-learning in nursing education: a systematic review and meta-analysis. *Nurse Educ Today* 2017;50:97-103. [doi: [10.1016/j.nedt.2016.12.020](https://doi.org/10.1016/j.nedt.2016.12.020)] [Medline: [28038371](#)]
98. Zhang H, Mörelius E, Goh SH, Wang W. Effectiveness of video-assisted debriefing in simulation-based health professions education: a systematic review of quantitative evidence. *Nurse Educ* 2019;44(3):E1-E6. [doi: [10.1097/NNE.0000000000000562](https://doi.org/10.1097/NNE.0000000000000562)] [Medline: [30015683](#)]
99. Qiao W, Bai Y, Lv R, Zhang W, Chen Y, Lei S, et al. The effect of virtual endoscopy simulator training on novices: a systematic review. *PLoS One* 2014;9(2):e89224 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0089224](https://doi.org/10.1371/journal.pone.0089224)] [Medline: [24586609](#)]
100. Berry MC, de M Neto JM, de Souza MI, Figueredo CM, Reher V, Evans JL. Effectiveness of technology-enhanced learning to improve periodontics educational outcomes: a systematic review. *J Dent Educ* 2020;84(7):830-839. [doi: [10.1002/jdd.12179](https://doi.org/10.1002/jdd.12179)] [Medline: [32421234](#)]
101. Lahti M, Hätönen H, Välimäki M. Impact of e-learning on nurses' and student nurses knowledge, skills, and satisfaction: a systematic review and meta-analysis. *Int J Nurs Stud* 2014;51(1):136-149. [doi: [10.1016/j.ijnurstu.2012.12.017](https://doi.org/10.1016/j.ijnurstu.2012.12.017)] [Medline: [23384695](#)]

102. Guedes HG, Câmara Costa Ferreira ZM, Ribeiro de Sousa Leão L, Souza Montero EF, Otoch JP, Artifon EL. Virtual reality simulator versus box-trainer to teach minimally invasive procedures: a meta-analysis. *Int J Surg* 2019;61:60-68 [FREE Full text] [doi: [10.1016/j.ijssu.2018.12.001](https://doi.org/10.1016/j.ijssu.2018.12.001)] [Medline: [30529216](https://pubmed.ncbi.nlm.nih.gov/30529216/)]
103. Kim JH, Park H. Effects of smartphone-based mobile learning in nursing education: a systematic review and meta-analysis. *Asian Nurs Res (Korean Soc Nurs Sci)* 2019;13(1):20-29 [FREE Full text] [doi: [10.1016/j.anr.2019.01.005](https://doi.org/10.1016/j.anr.2019.01.005)] [Medline: [30659927](https://pubmed.ncbi.nlm.nih.gov/30659927/)]
104. Gorbanev I, Agudelo-Londoño S, González RA, Cortes A, Pomares A, Delgadillo V, et al. A systematic review of serious games in medical education: quality of evidence and pedagogical strategy. *Med Educ Online* 2018;23(1):1438718 [FREE Full text] [doi: [10.1080/10872981.2018.1438718](https://doi.org/10.1080/10872981.2018.1438718)] [Medline: [29457760](https://pubmed.ncbi.nlm.nih.gov/29457760/)]
105. Tarpada SP, Morris MT, Burton DA. E-learning in orthopedic surgery training: a systematic review. *J Orthop* 2016;13(4):425-430 [FREE Full text] [doi: [10.1016/j.jor.2016.09.004](https://doi.org/10.1016/j.jor.2016.09.004)] [Medline: [27688638](https://pubmed.ncbi.nlm.nih.gov/27688638/)]
106. Kato C, Kataoka Y. Literature review of simulation training for acute obstetric emergencies. *J Jpn Acad Midwife* 2015;29(1):4-14. [doi: [10.3418/jjam.29.4](https://doi.org/10.3418/jjam.29.4)]
107. Zhou Y, Yang Y, Liu L, Zeng Z. [Effectiveness of mobile learning in medical education: a systematic review]. *Nan Fang Yi Ke Da Xue Xue Bao* 2018;38(11):1395-1400 [FREE Full text] [doi: [10.12122/j.issn.1673-4254.2018.11.20](https://doi.org/10.12122/j.issn.1673-4254.2018.11.20)] [Medline: [30514692](https://pubmed.ncbi.nlm.nih.gov/30514692/)]
108. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials* 1996;17(1):1-12. [doi: [10.1016/0197-2456\(95\)00134-4](https://doi.org/10.1016/0197-2456(95)00134-4)] [Medline: [8721797](https://pubmed.ncbi.nlm.nih.gov/8721797/)]
109. Li C, He J, Yuan C, Chen B, Sun Z. The effects of blended learning on knowledge, skills, and satisfaction in nursing students: a meta-analysis. *Nurse Educ Today* 2019;82:51-57. [doi: [10.1016/j.nedt.2019.08.004](https://doi.org/10.1016/j.nedt.2019.08.004)] [Medline: [31437783](https://pubmed.ncbi.nlm.nih.gov/31437783/)]
110. Higgins JP, 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;343:d5928 [FREE Full text] [doi: [10.1136/bmj.d5928](https://doi.org/10.1136/bmj.d5928)] [Medline: [22008217](https://pubmed.ncbi.nlm.nih.gov/22008217/)]
111. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64(4):383-394. [doi: [10.1016/j.jclinepi.2010.04.026](https://doi.org/10.1016/j.jclinepi.2010.04.026)] [Medline: [21195583](https://pubmed.ncbi.nlm.nih.gov/21195583/)]
112. Reed DA, Cook DA, Beckman TJ, Levine RB, Kern DE, Wright SM. Association between funding and quality of published medical education research. *JAMA* 2007;298(9):1002-1009. [doi: [10.1001/jama.298.9.1002](https://doi.org/10.1001/jama.298.9.1002)] [Medline: [17785645](https://pubmed.ncbi.nlm.nih.gov/17785645/)]
113. Ali S, Uppal MA, Gulliver SR. A conceptual framework highlighting e-learning implementation barriers. *Inf Technol People* 2018;31(1):156-180. [doi: [10.1108/itp-10-2016-0246](https://doi.org/10.1108/itp-10-2016-0246)]
114. Kituyi GM, Tusbira I. A framework for the integration of e-learning in higher education institutions in developing countries. *Int J Educ Dev Using Inf Commun Technol* 2013;9(2):19-36.
115. Madar MJ, Willis O. Strategic model of implementing e-learning. *Int J Sci Technol Res* 2014;3(5):235-238.
116. Masoumi D, Lindström B. Quality in e-learning: a framework for promoting and assuring quality in virtual institutions. *J Comput Assist Learn* 2012;28(1):27-41. [doi: [10.1111/j.1365-2729.2011.00440.x](https://doi.org/10.1111/j.1365-2729.2011.00440.x)]
117. Newton D, Ellis A. An e-learning comparative alignment framework. In: *World Conference on E-Learning in Corporate, Government, Healthcare & Higher Education*. 2007 Presented at: AACE '07; October 15-19, 2007; Chesapeake, VA.
118. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. *Acad Med* 2005;80(6):541-548. [doi: [10.1097/00001888-200506000-00005](https://doi.org/10.1097/00001888-200506000-00005)] [Medline: [15917356](https://pubmed.ncbi.nlm.nih.gov/15917356/)]
119. Ruiz JG, Cook DA, Levinson AJ. Computer animations in medical education: a critical literature review. *Med Educ* 2009;43(9):838-846. [doi: [10.1111/j.1365-2923.2009.03429.x](https://doi.org/10.1111/j.1365-2923.2009.03429.x)] [Medline: [19709008](https://pubmed.ncbi.nlm.nih.gov/19709008/)]
120. Cook DA, Beckman TJ. Reflections on experimental research in medical education. *Adv Health Sci Educ Theory Pract* 2010;15(3):455-464. [doi: [10.1007/s10459-008-9117-3](https://doi.org/10.1007/s10459-008-9117-3)] [Medline: [18427941](https://pubmed.ncbi.nlm.nih.gov/18427941/)]

Abbreviations

VR: virtual reality

Edited by G Eysenbach; submitted 16.07.21; peer-reviewed by R Bajpai, A Ardekani; comments to author 07.10.21; revised version received 07.11.21; accepted 21.11.21; published 17.03.22.

Please cite as:

Tudor Car L, Poon S, Kyaw BM, Cook DA, Ward V, Atun R, Majeed A, Johnston J, van der Kleij RMJJ, Molokhia M, V Wangenheim F, Lupton M, Chavannes N, Ajuebor O, Prober CG, Car J

Digital Education for Health Professionals: An Evidence Map, Conceptual Framework, and Research Agenda

J Med Internet Res 2022;24(3):e31977

URL: <https://www.jmir.org/2022/3/e31977>

doi: [10.2196/31977](https://doi.org/10.2196/31977)

PMID: [35297767](https://pubmed.ncbi.nlm.nih.gov/35297767/)

©Lorraine Tudor Car, Selina Poon, Bhone Myint Kyaw, David A Cook, Victoria Ward, Rifat Atun, Azeem Majeed, Jamie Johnston, Rianne M J J van der Kleij, Mariam Molokhia, Florian V Wangenheim, Martin Lupton, Niels Chavannes, Onyema Ajuebor, Charles G Prober, Josip Car. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 17.03.2022. 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.

Original Paper

A Question-and-Answer System to Extract Data From Free-Text Oncological Pathology Reports (CancerBERT Network): Development Study

Joseph Ross Mitchell^{1,2,3}, MSc, PhD; Phillip Szepietowski⁴, PhD; Rachel Howard⁴, PhD; Phillip Reisman⁴, MS; Jennie D Jones⁴, MSHI-HA; Patricia Lewis⁴, BSc; Brooke L Fridley⁵, PhD; Dana E Rollison⁴, PhD

¹Department of Machine Learning, H Lee Moffitt Cancer Center and Research Institute, Tampa, FL, United States

²Department of Medicine, Faculty of Medicine & Dentistry, and the Alberta Machine Intelligence Institute, University of Alberta, Edmonton, AB, Canada

³Alberta Health Services, Edmonton, AB, Canada

⁴Department of Health Data Services, H Lee Moffitt Cancer Center and Research Institute, Tampa, FL, United States

⁵Department of Biostatistics and Bioinformatics, H Lee Moffitt Cancer Center and Research Institute, Tampa, FL, United States

Corresponding Author:

Dana E Rollison, PhD

Department of Health Data Services

H Lee Moffitt Cancer Center and Research Institute

12902 Magnolia Drive

Tampa, FL, 33612

United States

Phone: 1 813 745 6530

Email: Dana.Rollison@moffitt.org

Abstract

Background: Information in pathology reports is critical for cancer care. Natural language processing (NLP) systems used to extract information from pathology reports are often narrow in scope or require extensive tuning. Consequently, there is growing interest in automated deep learning approaches. A powerful new NLP algorithm, bidirectional encoder representations from transformers (BERT), was published in late 2018. BERT set new performance standards on tasks as diverse as question answering, named entity recognition, speech recognition, and more.

Objective: The aim of this study is to develop a BERT-based system to automatically extract detailed tumor site and histology information from free-text oncological pathology reports.

Methods: We pursued three specific aims: extract accurate tumor site and histology descriptions from free-text pathology reports, accommodate the diverse terminology used to indicate the same pathology, and provide accurate standardized tumor site and histology codes for use by downstream applications. We first trained a base language model to comprehend the technical language in pathology reports. This involved unsupervised learning on a training corpus of 275,605 electronic pathology reports from 164,531 unique patients that included 121 million words. Next, we trained a question-and-answer (Q&A) model that connects a Q&A layer to the base pathology language model to answer pathology questions. Our Q&A system was designed to search for the answers to two predefined questions in each pathology report: *What organ contains the tumor?* and *What is the kind of tumor or carcinoma?* This involved supervised training on 8197 pathology reports, each with ground truth answers to these 2 questions determined by certified tumor registrars. The data set included 214 tumor sites and 193 histologies. The tumor site and histology phrases extracted by the Q&A model were used to predict International Classification of Diseases for Oncology, Third Edition (ICD-O-3), site and histology codes. This involved fine-tuning two additional BERT models: one to predict site codes and another to predict histology codes. Our final system includes a network of 3 BERT-based models. We call this CancerBERT network (caBERTnet). We evaluated caBERTnet using a sequestered test data set of 2050 pathology reports with ground truth answers determined by certified tumor registrars.

Results: caBERTnet's accuracies for predicting group-level site and histology codes were 93.53% (1895/2026) and 97.6% (1993/2042), respectively. The top 5 accuracies for predicting fine-grained ICD-O-3 site and histology codes with 5 or more samples each in the training data set were 92.95% (1794/1930) and 96.01% (1853/1930), respectively.

Conclusions: We have developed an NLP system that outperforms existing algorithms at predicting ICD-O-3 codes across an extensive range of tumor sites and histologies. Our new system could help reduce treatment delays, increase enrollment in clinical trials of new therapies, and improve patient outcomes.

(*J Med Internet Res* 2022;24(3):e27210) doi:[10.2196/27210](https://doi.org/10.2196/27210)

KEYWORDS

natural language processing; NLP; BERT; transformer; pathology; ICD-O-3; deep learning; cancer

Introduction

Background

Much of the information in electronic medical records (EMRs) required for the practice of clinical oncology and cancer research is contained in unstructured text. As much as 80% of EMR data can be found in narrative text notes and scanned documents [1], ranging from clinic or surgical notes, including pathology, radiology, and ambulatory care, to past medical or family history. The extraction of discrete data elements from this unstructured text, particularly those relating to disease diagnosis and commonly captured in routine pathology reports, is critical for the selection of treatment options, identification of patients eligible for clinical trials, and monitoring of adherence to established clinical treatment pathways.

Natural language processing (NLP) has been used to extract information from medical text for several decades [2-6], and a thorough review of NLP-based information extraction for cancer-related EMR notes can be found in the study by Datta et al [7]. Application of NLP to pathology reports has also been prevalent in the literature during the course of the last decade [8-19]. That said, many early studies used regular expression- and rule-based systems [20,21] that require considerable up-front development and can be difficult to adapt and maintain.

More recently, there has been growing interest in more highly automated deep learning approaches for clinical NLP [6,22]. In late 2018, a powerful new deep learning NLP algorithm was released: bidirectional encoder representations from transformers (BERT) [23]. BERT established new, state-of-the-art performance levels on common nonclinical NLP benchmarks [24]. This success spawned rapid research and development of multiple BERT-inspired and transformer-based neural architectures [25-33]. Several of these have, for the first time, achieved or surpassed human-level performance on tasks as diverse as question answering, named entity recognition, speech recognition, and more [30,33-35]. BERT and related architectures have facilitated significant improvements in multiple medical applications, including processing of electronic health records [36,37], outcome prediction [38-40], identification of medical terms and concepts [41], medical chatbots [42], sentiment analysis [43], recommender systems [44], and others. Despite BERT's success, we are aware of only a single application of BERT to the already promising area of free-text pathology reports [45]. The study focused on classification of text into only a few cancer-related categories, including afflicted organ (15 organ groups), disease type (noncancer, premalignant, or cancer), cancer reason (6 histology groups), and presence of metastatic disease (no, yes: in lymph nodes and yes: in non-lymph node tissue). Our goal is to develop and evaluate a

BERT-based system to extract detailed tumor site and histology information from free-text pathology reports. The availability of manually curated data within the H Lee Moffitt Cancer Center and Research Institute (Moffitt) Cancer Registry (MCR) represented a unique opportunity to train a BERT-based system using a gold standard data set classified using a standard ontology.

BERT's proficiency at question answering prompted us to construct a question-and-answer (Q&A) system to extract clinical data from pathology reports. This concept has long been compelling—Q&A systems for medical data extraction have been pursued for >40 years [46]. Such a system would have several desirable properties: an intuitive user interface, the ability to extract additional data fields by searching for answers to additional questions, and the ability to generalize to other medical documents. Furthermore, it would allow us to make data available for clinical and research use close to real time, thus reducing treatment delays, increasing enrollment in clinical trials of new therapies, and improving patient outcomes. To train such a general-purpose Q&A system on pathology reports, one would need a diverse set of questions on which to train it. Our task in this paper is more modest (and is in essence a classification task of site and histology); however, we view the Q&A portion of our system as a small step toward this broader goal.

Primary Contributions

This work describes three primary contributions:

1. A new BERT language model for comprehension of pathology reports in oncology. We call this *CancerBERT*, or *caBERT* for short.
2. A new Q&A caBERT-based system, tolerant to varied terminologies, word orders, and spelling mistakes, to extract tumor site and histology descriptions from free-text pathology reports.
3. A new caBERT network (caBERTnet) to predict International Classification of Diseases for Oncology, version 3.2 (ICD-O-3.2), codes from the extracted descriptions. This system can handle up to 332 organ sites and 1143 tumor histologies. On an unseen test data set with 214 sites and 193 histologies it achieved overall accuracies that are equal to, or above, those of existing systems, while also expanding on the number of site and histology classes captured by these systems. Although the results in practice still require human validation, they provide a means of early abstraction from unstructured pathology text over a very broad set of sites and histologies and in addition can provide an initial signal to assist expert cancer registrars in their case diagnosis-abstraction workflow.

Methods

Information on our software development tools is provided in [Multimedia Appendix 1 \[47-53\]](#). To construct our system, we had to achieve three specific aims: (1) extract accurate tumor site and histology descriptions from complex free-text pathology reports, (2) accommodate the diverse terminology used to indicate the same pathology, and (3) provide accurate standardized tumor site and histology codes for use by downstream applications.

Textbox 1. A fragment of text from a pathology report generated at H Lee Moffitt Cancer Center and Research Institute (Moffitt).

Fragment of text from a pathology report generated at Moffitt

- clinical history: not given. preoperative diagnosis: right lower lobe, squamous cell carcinoma. specimen types: a: right station 7 fs b: station # 10 c: right lung fs d: station 4r e: additional station 10 f: additional station 7 final diagnosis: a. lymph node right station 7 biopsy: anthracotic lymph node 1 with lymphoid hyperplasia negative for malignancy. b. lymph node station 10 biopsy: anthracotic lymph node 1 negative for malignancy. c. lung right pneumonectomy: moderate to poorly differentiated squamous cell carcinoma with extensive necrosis see key pathological findings. bronchial and vascular resection margins negative for malignancy. three of 10 hilar lymph nodes with metastatic squamous cell carcinoma. d. lymph nodes station 4r biopsies: anthracotic lymph nodes 5 negative for malignancy. e. lymph node additional station 10 biopsy: minute lymph node 1 negative for malignancy. f. lymph node station 7 biopsy: anthracotic lymph nodes 4 negative for malignancy. key pathological findings tumor type: squamous cell carcinoma with extensive necrosis. histological grade: moderate to poorly differentiated. tumor location: right lung involving right lower lobe. right upper and middle lobes free of tumor.

Next, it required us to train a Q&A model that appends a Q&A layer onto the pathology language model to answer pathology questions. Our Q&A system was designed to search for the answers to two predefined questions in each pathology report:

1. What organ contains the tumor?
2. What is the kind of tumor or carcinoma?

For example, when presented with the report shown in [Textbox 1](#), the system should respond to the question *What organ contains the tumor?* with *C343: lower lobe, lung*, and would respond to the question *What is the kind of tumor or carcinoma?* with *8070/3: squamous cell carcinoma, nos* (not otherwise specified).

This involved *supervised* training on a set of pathology reports, each with ground truth answers to these 2 questions determined by human experts. To do this, we constructed a second *fine-tuning training data set*, described in more detail in the

Extract Tumor Site and Histology Descriptions

Overview

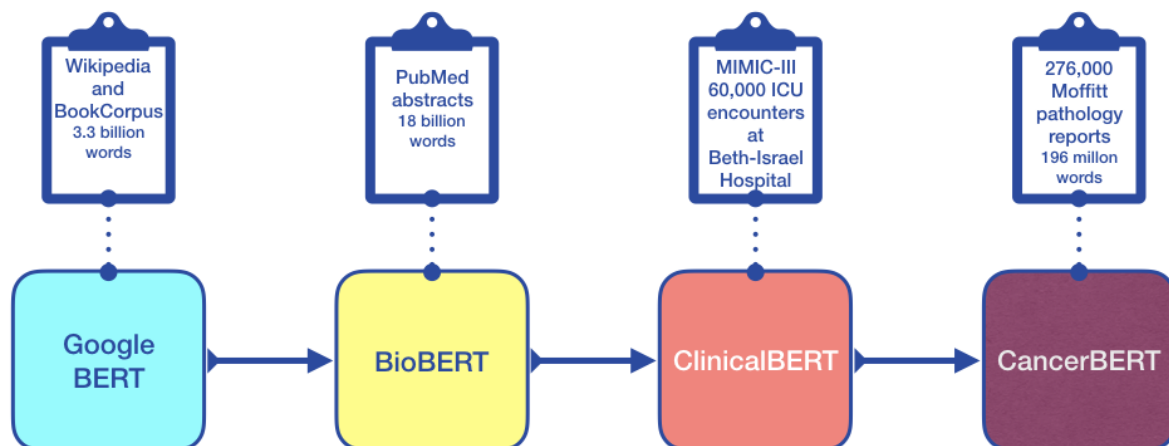
Constructing this system required us to first train a base language model to comprehend the technical language in pathology reports ([Textbox 1](#)). This involved *unsupervised* training on a large corpus of pathology text. For this we constructed a *pathology language-model training data set*, described in more detail in the next section (*Pathology Language Model*).

Pathology Q&A Model section. Finally, we evaluated our system using a sequestered *fine-tuning testing data set*, described in more detail in the *Pathology Q&A Model* section.

Pathology Language Model

Training a base language model to comprehend pathology reports leveraged prior work by several groups ([Figure 1](#)). Lee et al [54] performed transfer learning on BERT using nearly 18 billion words extracted from PubMed abstracts. The result, BioBERT, is tuned for biomedical language comprehension tasks and is publicly available. Next, Alsentzer et al [55] performed transfer learning on BioBERT to tune it for clinical language comprehension. They used EMR notes in the Medical Information Mart for Intensive Care, version 3 (MIMIC-III) data set [56], which includes data from approximately 60,000 intensive care unit stays by patients at Beth-Israel Hospital in Boston, Massachusetts. The model created by Alsentzer et al [55], ClinicalBERT, was also made publicly available.

Figure 1. Sequence of transfer-learning steps used in training the CancerBERT base language model. BERT: bidirectional encoder representations from transformers; ICU: intensive care unit; MIMIC-III: Medical Information Mart for Intensive Care, version 3.



Alsentzer et al [55] created two models built upon BioBERT: one trained on all MIMIC-III notes and one trained on just the MIMIC-III discharge summaries. Initial pretraining experimentation revealed that the latter provided higher accuracies on a separate sample of our pathology reports. We noted that Moffitt pathology reports have a language structure closer to discharge summaries than to general clinical notes. Consequently, our model was initialized with weights from the latter of the two ClinicalBERT models: ClinicalBERT–Bio+Discharge Summary BERT Model.

Transfer learning was accomplished by performing masked-language modeling [23]. Briefly, 15% of the words in the corpus are selected at random and replaced with a *mask* token. The language model is then trained to predict the masked words. The word-masking process is performed automatically at the beginning of each training run.

Our language-model training corpus included electronic pathology reports of solid tumors produced by pathologists at Moffitt between 1986 and 2020. The year 1986 was the earliest date on pathology reports cataloged in our enterprise data

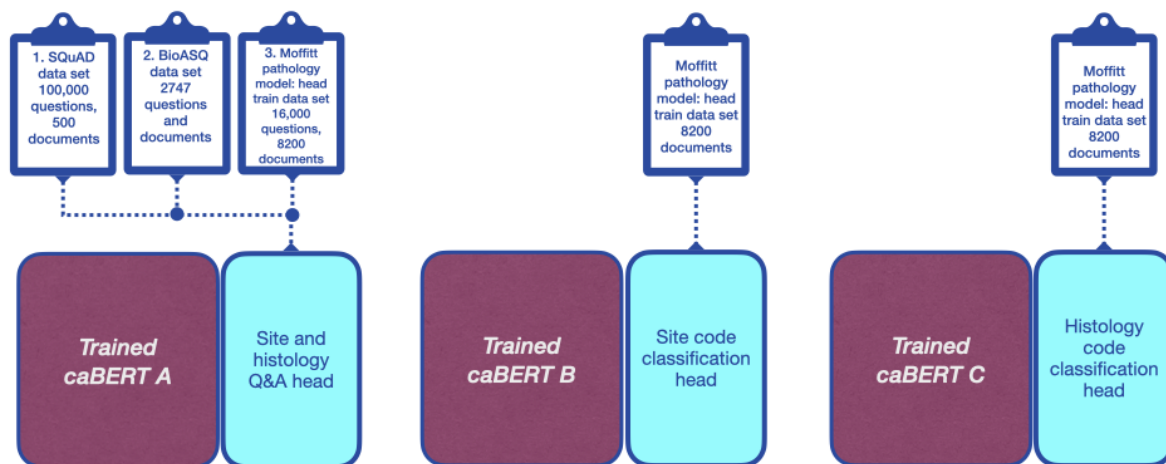
warehouse. The data set was restricted to solid tumors for two reasons: first, to focus the problem domain for this proof-of-concept study, and, second, Moffitt hematologic pathology reports follow a quasi-structured format, reducing the need for extraction of data from unstructured text.

This data set contained both Health Level Seven International messages and plain-text pathology reports. These were minimally processed to extract and clean the text relevant to pathology (more details can be found in [Multimedia Appendix 1](#)). The final language model training corpus included 275,605 electronic pathology reports from 164,531 unique patients and included 121 million words.

Pathology Q&A Model

The pathology Q&A lesson plan involved 3 stages, each intended to improve our system’s comprehension of pathology reports and thereby increase the accuracy of question answering ([Figure 2](#)). The three stages involved training the Q&A model to (1) answer general English language questions, (2) answer technical biomedical science questions, and (3) answer questions from Moffitt pathology reports.

Figure 2. Lesson plan for the caBERT network consisting of one question and answering model A and two classification models, one for primary site (model B) and another for histology (model C). BioASQ: Biomedical Semantic Indexing and Question Answering; Q&A: question and answer; SQuAD: Stanford Question Answering Dataset.



Each training stage used supervised learning. This required a training data set that included passages of text, ≥ 1 questions related to each passage, and ground truth answers to those questions that appeared as contiguous phrases within the related passage. At the end of each stage we evaluated our system using the same sequestered test data set constructed from Moffitt pathology reports, as described in more detail later in this section. The experimental parameters used to train the Q&A model were held constant over all stages and are listed in [Table S1](#) in [Multimedia Appendix 1](#).

For the first stage of training we used the Stanford Question Answering Dataset (SQuAD), version 1.1 [57]. SQuAD consists of more than 100,000 questions and answers created by crowdworkers on Wikipedia articles. The SQuAD data format is widely used in NLP research. Therefore, we designed our system to read and process data sets in this format.

For the second stage of training we used the large-scale Biomedical Semantic Indexing and Question Answering (BioASQ) data set [58]. In particular, we used data from BioASQ Challenge 7b: Biomedical Semantic Question Answering. This data set contains 2747 training questions along with their ground truth answers. According to the BioASQ Challenge 7b description: “All the questions are constructed by biomedical experts from around Europe.” This data set was converted to SQuAD format by Yoon et al [59] and made available for public use.

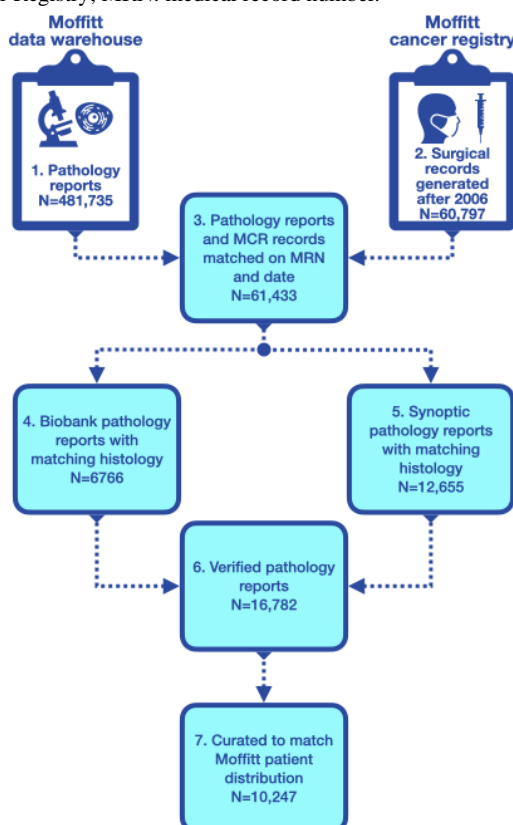
Next, we constructed a Q&A fine-tuning data set in SQuAD format based on Moffitt pathology reports. We obtained ground truth answers to our 2 questions from data abstracted by Moffitt certified tumor registrars (CTRs). CTRs undergo an extensive training and internship program to become proficient at extracting quantitative and categorical data from unstructured pathology reports. They are widely employed by cancer centers

and other organizations to extract data for clinical and research applications and for reporting to state and national agencies. The MCR deploys state-of-the-art quality assurance procedures: its benchmark for quality is 90% and its target accuracy is 95% [60].

Moffitt's enterprise data warehouse was searched to find solid tumor pathology reports generated after 2006 with matched MCR data (Figure 3). These reports were screened to ensure

that they contained a description of a positive diagnosis of a single primary tumor. Next, each report was processed to ensure that it contained an answer to at least one of the questions in the Q&A model. This was accomplished programmatically by searching each report for a phrase contained in a table of acceptable answer phrases (Figure S1 in Multimedia Appendix 1), as described in more detail in the next section (*Accommodate Diverse Terminology*). Our search produced 16782 reports that met these inclusion criteria (Figure 3).

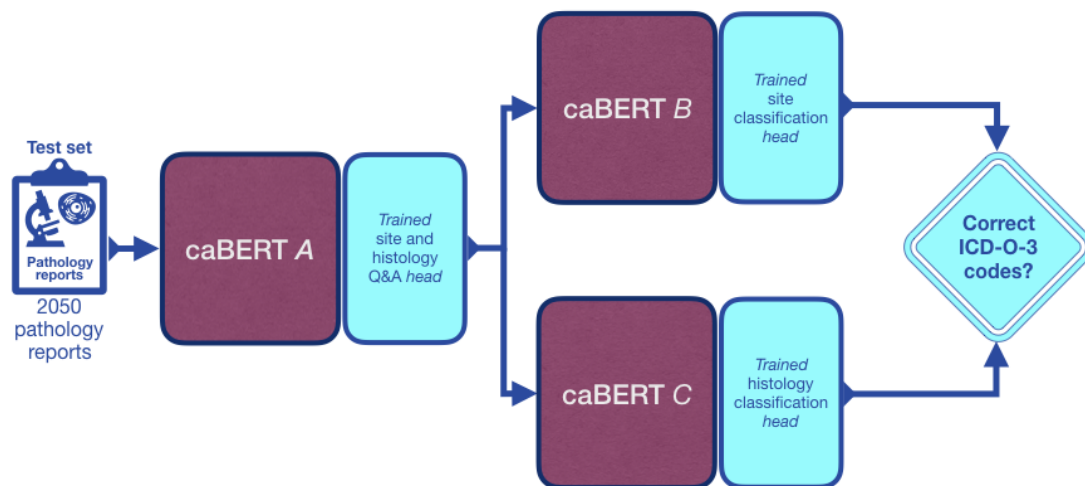
Figure 3. Flowchart depicting the data curation process for creating the Moffitt fine-tuning data sets (used in the site and histology question-and-answer and classification tasks). MCR: Moffitt Cancer Registry; MRN: medical record number.



Next, we curated these reports to ensure that (1) the relative frequencies of the 10 most common tumor sites and histologies in this collection matched the relative frequencies in Moffitt's patient population as a whole and (2) all MCR-assigned tumor sites and histologies reported for Moffitt patients after 2006 were represented in the data set. The final curated collection contained 10,247 reports (Figure 3). We will refer to this data set as the *fine-tuning* data set to contrast it with the caBERT language model data set used to train the base language model as described in the *Pathology Language Model* section.

The curated collection of reports was randomly divided to create two data sets: 79.99% (8197/10,247) of the reports were used to create the Moffitt fine-tuning training data set and 20.01% (2050/10,247) were used to create the Moffitt fine-tuning testing data set. Each data set was saved in SQuAD format using a custom-written Python program. The training data set was used for the final stage of Q&A training and also for ICD-O-3 code predictions, described in more detail in the *Accurate ICD-O-3 Codes* section. The testing subset was used to evaluate the impact of Q&A training at the end of each training stage (Figure S2 in Multimedia Appendix 1) and also to evaluate the performance of the final pipeline (Figure 4).

Figure 4. The final caBERT network (caBERTnet) connects caBERT instances A, B, and C, used for site and histology question and answering, International Classification of Diseases for Oncology, Third Edition (ICD-O-3) primary site and ICD-O-3 histology code classification, respectively.



Question-answering accuracy was evaluated using two metrics: exact match and F1 score. Exact match is true if the caBERT-extracted phrase is an identical word-for-word match with the MCR phrase and false otherwise. We calculated the average number of true results across all test samples. The F1 score is a measure of the degree of overlap among the words in the caBERT-extracted phrase and the MCR phrase. This varies from 0 (no words in common) to 1 (all words in common but not necessarily in the same order). Each exact match corresponded to an F1 score of 1.0. We calculated the average F1 score across all test samples and expressed it as a percentage.

After all training stages were complete, they were repeated using the initial ClinicalBERT model (*ClinicalBERT-Bio+Discharge Summary BERT*) with a new randomly initialized Q&A layer. This allowed us to determine the impact of developing a pathology-tuned BERT model on extraction accuracy over the baseline accuracy of using ClinicalBERT alone (Figure S2 in [Multimedia Appendix 1](#)). The training parameters were set to those optimized for ClinicalBERT and reported by Alsentzer et al [55].

Accommodate Diverse Terminology

Our pathology reports came supplied with ground truth labels for primary site and histology in the form of ICD-O-3 codes [61], which were abstracted by Moffitt CTRs. To train the Q&A model we needed a method to determine the precise location within each pathology report of the actual text corresponding to these labels. This proved nontrivial owing to the rather diverse terminology within each pathology report used to refer to each primary site and histology.

To address this issue, we used data from several canonical sources. Our primary source was the ICD-O-3 standards [62], which we used to define the primary *preferred* terminology for each code. Within the ICD-O-3 standards there are 332 unique site codes and 1143 unique histology codes, each with accompanying preferred terms. Along with the preferred term, many codes also have an additional set of synonyms, which we stored together with the preferred term in a table of acceptable phrases for each code. In addition to the ICD-O-3 tables, we also used terminology from the National Cancer Institute's

Surveillance, Epidemiology, and End Results (SEER) program Site/Histology Validation List [62], along with the SEER Site-specific training module website [63].

To provide a little more detail, the specific sources we used to construct our acceptable phrase tables were as follows. For histology, we used the ICD-O-3.2 morphology table (version 15112019) [64] and supplemented this with terms from the SEER Site/Histology Validation List (version 20150918), current versions of which are both available in Microsoft Excel format from their respective websites. For the site terms, we used the ICD-O-3 mapping table maintained by the National Cancer Institute [65], supplemented again by the SEER Site/Histology Validation List. In addition, for the site terms we also scraped the tables contained in the SEER Site-specific learning module website for any new terms.

The aforementioned sources have the benefit of being subject to an international standard and are useful in designating preferred terms for each histology and site code. However, we should note that there do exist slight discrepancies between the World Health Organization-maintained ICD-O-3 coding standards and the North American Association of Central Cancer Registries coding guidelines, which are followed in the SEER materials. For simplicity, we chose to base our model on the ICD-O-3 standards, but this caveat may prove relevant for any future cancer registry applications of the model.

Although these sources provided us with preferred and alternative terminologies, they did not encompass the full range of language used for every label in our pathology reports, which often included things such as permutations of word orderings as well as acronyms and other typographical differences with the canonical terms. Of note, Moffitt CTRs routinely record a short description of the histology and site for every labeled pathology report in a text-based field. For each histology and site code, we appended these additional phrases to the list of synonyms of the preferred canonical terminology.

Using the sources described earlier, we created two hierarchical tree structures as illustrated in Table S2 in [Multimedia Appendix 1](#): one to hold histology terms and one to hold site terms. To construct these trees, the histology and site codes were first

grouped into broad morphology and site groups as specified in the ICD-O-3 tables. Within each group are a collection of specific codes, where each code has an associated preferred term, and a list of synonyms. For efficient searching, these trees were stored as JSON objects that were imported into Python as nested dictionaries and lists. See Figure S1 in [Multimedia Appendix 1](#) for an example of an entry in our acceptable phrase table.

To search each pathology report for appropriate spans of text, we used the trees to construct a dictionary with keys provided by the specific site and histology codes and values provided by the associated acceptable phrase table. Using this dictionary, for each pathology report we implemented a simple search for an exact match from the list of preferred terms and synonyms for the labeled ground truth histology and site code, giving preference for the preferred term, followed by each synonym ordered by length (with the longest matching synonym given preference over the others).

Even with the diverse terminology within the acceptable phrase table for each code, not every pathology report contained an exact match within the list of allowed terms. For pathology reports that did not contain an exact match, we further refined the search by allowing for matches that only overlapped with a subset of the word tokens within each phrase, again giving preference to the longest synonyms and also using a set of stop terms to avoid overly general terminology. To capture potential word-ordering differences, we allowed these word token subsets to be constructed in an arbitrarily permuted order, which was made efficient by using the *itertools* module available as part of the Python Standard Library.

Using the aforementioned procedure, of the 10,247 pathology reports in the Moffitt fine-tuning training and testing data sets, we were able to find appropriate textual answers within 10,096 (98.53%) reports for primary site (with $n=8070$, 79.93%, in the training set and $n=2026$, 20.07%, in the testing set) and within 10,218 (99.72%) reports for histology (with $n=8176$, 80.02%, in the training set and $n=2042$, 19.98%, in the testing set).

Provide Accurate ICD-O-3 Codes

Overview

The tumor site and histology phrases extracted by the Q&A model ([Figure 2](#)) were used to predict ICD-O-3 site and histology codes. This involved fine-tuning two additional copies of caBERT: one to predict site codes and the second to predict histology codes. Our final system ([Figure 4](#)) includes a network of 3 caBERT-based models. We refer to this system as caBERTnet.

Training the ICD-O-3 Site and Histology Code Classifiers

Classifier training parameters are described in more detail in Table S1 in [Multimedia Appendix 1](#). Briefly, each caBERT instance was trained to perform a classification task: given an input phrase, predict the corresponding ICD-O-3 code. Classification tasks were trained using the Moffitt fine-tuning training data set (*Extract Tumor Site and Histology Descriptions* section). Training samples were screened to ensure that each

contained ground truth site and histology codes and at least one site or histology phrase provided by the MCR. Missing site and histology phrases were filled using SEER preferred terms. These were identified by performing a lookup into the ICD-O-3 table using the site or histology code in the training sample.

After screening, the ground truth phrases were labeled and concatenated to form a single combined phrase. For example, if the MCR phrases were *lung lower lobe* and *squamous cell carcinoma*, then the combined phrase would be *site: lung lower lobe. histology: squamous cell carcinoma*. The combined phrase was used to train both the caBERT site classification model and the caBERT histology classification model. The use of a combined phrase allowed caBERTnet to leverage any correlation between site and histology to improve its performance. For example, astrocytomas are brain tumors. When caBERTnet encountered a previously unseen pathology report during the test phase with the combined phrase *site: frontal. histology: anaplastic astrocytoma*, it correctly predicted a brain site of *C711, frontal lobe* and a histology of *9401/3 astrocytoma anaplastic NOS* (not otherwise specified).

Testing the ICD-O-3 Site and Histology Code Classifiers

After training of the site and histology ICD-O-3 code classification models was complete, caBERTnet performance was evaluated using the sequestered Moffitt fine-tuning test data set described in the *Extract Tumor Site and Histology Descriptions* section. For each test sample, the MCR-generated site and histology phrases were used to create a *ground truth* combined phrase. Next, the site and histology phrases extracted by the Q&A stage of caBERTnet were used to create a *predicted* combined phrase. The predicted phrase was tokenized to prepare it for input into each classification model. Ground truth site and histology codes from the MCR were enumerated, as described earlier, and stored as true labels. Subsequently, the trained site and histology classification models were used to classify the tokenized *predicted* combined phrases for each test sample. The outputs from this classification, logits, were converted into probabilities, sorted, and converted back into ICD-O-3 codes as described earlier, labeled as *predicted* codes, and saved for further performance analysis.

caBERTnet performance was evaluated in 3 different ways. First, the top 5 accuracies were determined. This metric (or its inverse, the top 5 error rate) is commonly used to evaluate classification algorithms [66]. Briefly, it calculates the average probability that the correct site or histology code occurs within the top N predicted codes because N is varied from 1 through 5. Top 1 accuracy, the accuracy of the code scored most highly by the classification algorithm, is equivalent to precision, recall, and F1 score for this classification task.

Second, we examined the effect of culling or removing infrequently occurring codes. Our hypothesis was that the caBERT site and histology code classifiers suffer when they do not have enough training data to learn from. Therefore, to examine the effect of training sample size, we iteratively eliminated site and histology codes from the full 2050-sample test data set when the number of examples with a particular code in the training set (alone) fell below a specified threshold. We varied that threshold from 0 samples (no culling) to 35

samples in increments of 5 samples. At each culling threshold we recalculated the top 5 performance of the site and histology classifiers.

Third, we calculated the overall accuracy of predicting the correct *group* code for each site and histology code. *Group* codes occur higher up in the ontological tree and, as the name implies, encompass a group or range of related tumor sites or histologies. For example, the site codes *C341 upper lobe, lung* and *C349 lung, NOS* have the same group code: *C34 bronchus and lung*. The histology codes *8070/3 squamous cell carcinoma, NOS* and *8051/0 verrucous carcinoma* both have the same group code: *805-808 squamous cell neoplasms*. The ICD-O-3 ontology includes 82 group-level site codes covering the 332 fine-grained site codes. It includes 49 group-level histology codes covering the 1143 fine-grained histology codes.

Group codes are useful for search and summary applications. The group codes for both the predicted and ground truth fine-grained codes were determined by searching in the tree data structures described in the *Accommodate Diverse Terminology* section. For each fine-grained code, the search started at that code’s location in the tree and proceeded upward. Finally, we calculated the overall accuracy of prediction within each group code for both site and histology predictions.

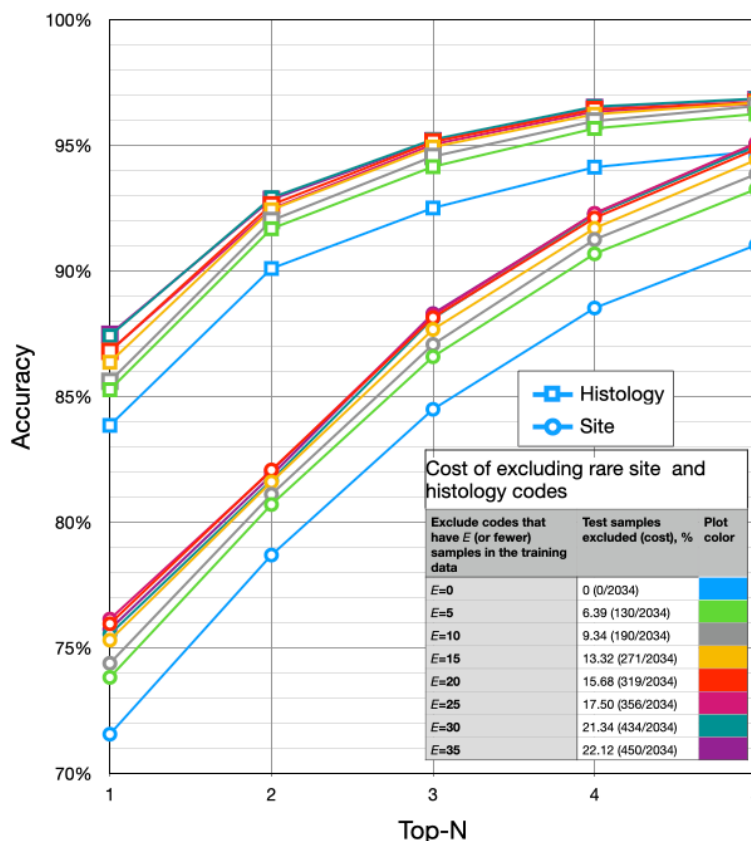
Results

Model Accuracy

The accuracy of the both the ClinicalBERT and caBERT Q&A models when tested on the Moffitt fine-tuning testing data set improved at each Q&A training stage (SQuAD, BioASQ, and Moffitt training data; Figure S2 in [Multimedia Appendix 1](#)). ClinicalBERT had higher accuracy than caBERT on the Moffitt test set after each of the first 2 training stages. This suggests that the specialized pathology-language tuning reduced caBERT’s ability to learn from the SQuAD and BioASQ training data sets. However, caBERT outperformed ClinicalBERT after training on Moffitt pathology reports. This was true both for exact match (3254/4068, 79.99%, for caBERT vs 3069/4068, 75.44%, for ClinicalBERT) and F1 score (87.76% for caBERT vs 84.85% for ClinicalBERT).

The top N accuracy of predicting fine-grained site codes ranged from 71.58% (1456/2034; top 1) to 91.05% (1852/2034; top 5), without culling (Figure 5). The accuracy for predicting histology codes ranged from 83.87% (1706/2034; top 1) to 94.79% (1928/2034; top 5). Culling 6.39% (130/2034) of the test samples—those site and histology codes with <5 samples in the training data set—improved accuracy for site code prediction to 73.84% (1406/1904; +2.26%; top 1) and 93.28% (1776/1904; +2.23%; top 5). The same culling improved the accuracy of histology code prediction to 85.29% (1624/1904; +1.42%; top 1) and 96.27% (1833/1904; +1.48%; top 5).

Figure 5. The effect of culling rare tumor sites and histologies on the top N accuracy of predicting fine-grained International Classification of Diseases for Oncology, Third Edition codes.

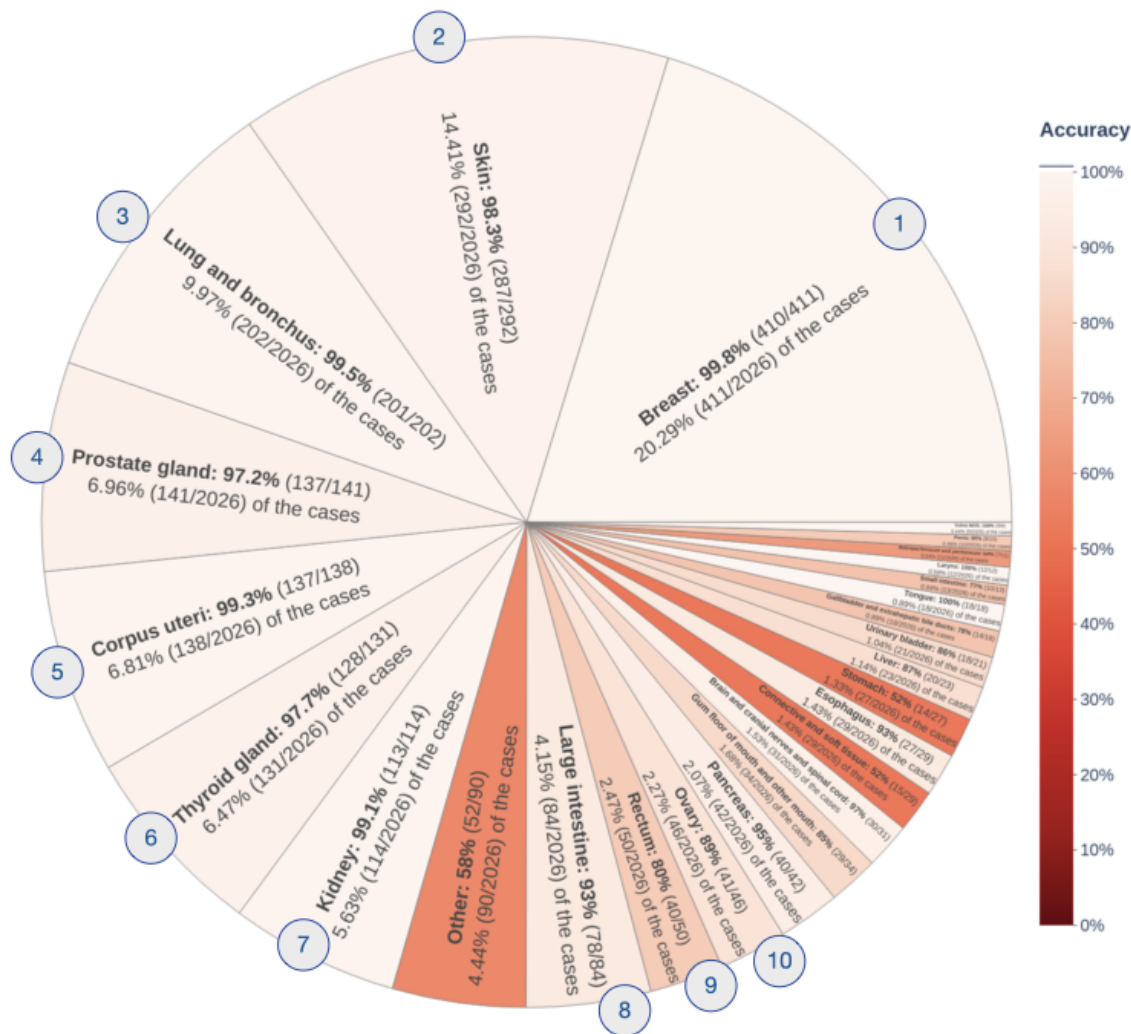


Group-Level Code Predictions

We also computed the accuracy of the more coarse-grained group-level code predictions by mapping each top 1 code prediction to its corresponding group in the ICD-O-3 ontology. The accuracy of predicting group-level site codes was 93.53% (1895/2026) overall (Figure 6). The ten most commonly represented sites—(1) breast, (2) skin, (3) lung and bronchus, (4) prostate gland, (5) corpus uteri, (6) thyroid gland, (7) kidney, (8) large intestine, (9) rectum, and (10) ovary—included 79.42%

(1609/2026) of the test samples and had an accuracy of 97.7% (1572/1609). Accuracies <80% were observed for connective and soft tissues (15/29, 52%, with 29/2026, 1.43%, of the samples), stomach (14/27, 52%, with 27/2026, 1.33%, of the samples), gallbladder and extrahepatic bile ducts (14/18, 78%, with 18/2026, 0.89%, of the samples), small intestine (10/13, 77%, with 13/2026, 0.64%, of the samples), retroperitoneum and peritoneum (7/11, 64%, with 11/2026, 0.54%, of the samples), and other (52/90, 58%, a collection of 27 sites totaling 90/2026, 4.44%, of the samples).

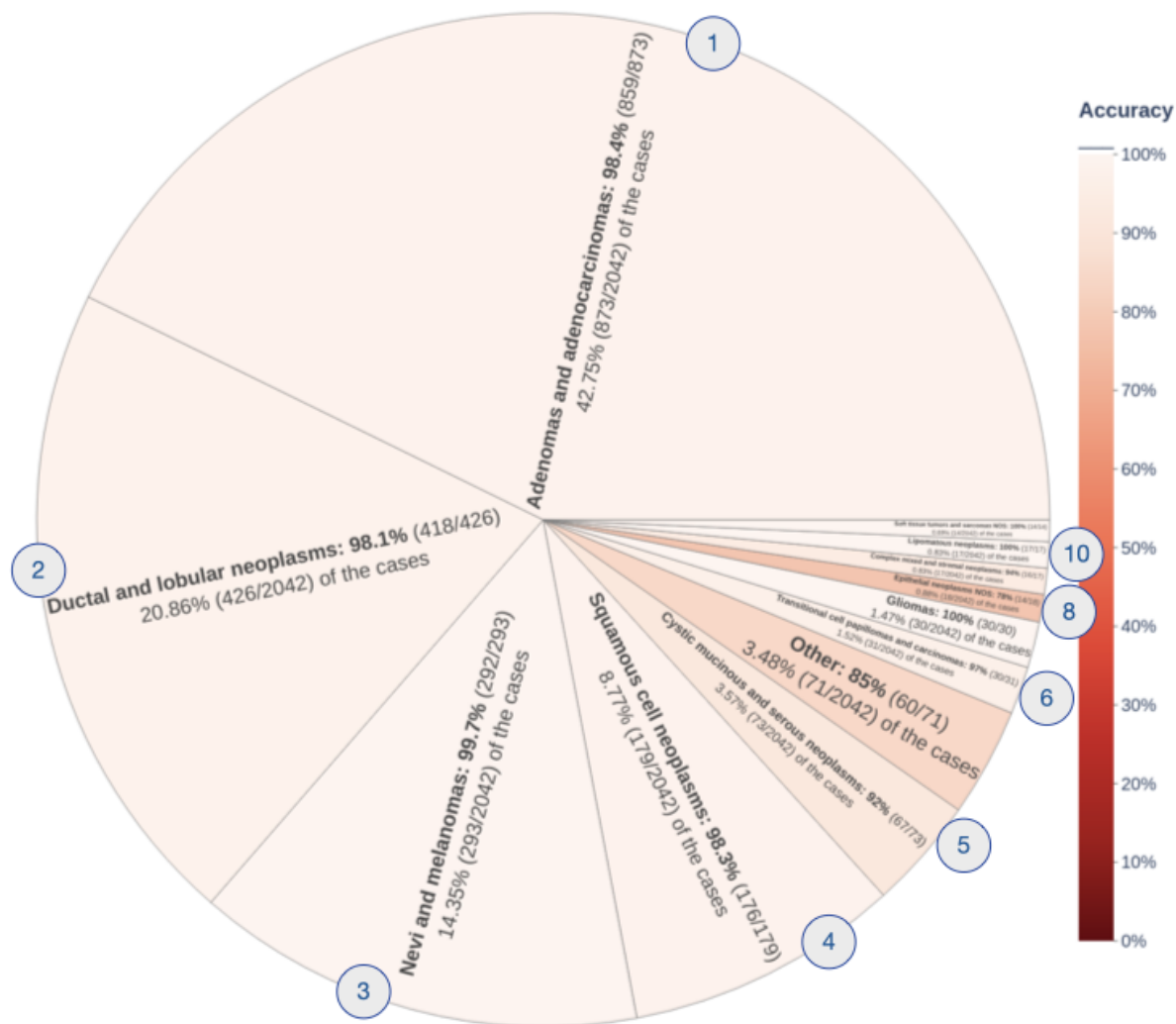
Figure 6. Accuracy of predicting tumor site group codes from unstructured and previously unseen pathology reports on solid tumors, broken down to show performance within each site group. The overall accuracy over all site groups was 93.53% (1895/2026).



The accuracy of predicting group-level histology codes was 97.6% (1993/2026) overall (Figure 7). The ten most commonly represented histologies—(1) adenomas and adenocarcinomas, (2) ductal and lobular neoplasms, (3) nevi and melanomas, (4) squamous cell carcinomas, (5) cystic mucinous and serous neoplasms, (6) transitional cell papillomas and carcinomas, (7)

gliomas, (8) epithelial neoplasms, (9) complex mixed and stromal neoplasms, and (10) lipomatous neoplasms—included 95.84% (1957/2042) of the test samples and had an accuracy of 98.06% (1919/1957). An accuracy of <80% was observed for epithelial neoplasms only (14/18, 78%, with 18/2042, 0.88%, of the samples).

Figure 7. Accuracy of predicting tumor histology group codes from unstructured and previously unseen pathology reports on solid tumors, broken down to show performance within each histology group. The overall accuracy over all histology groups was 97.6% (1993/2042).



Discussion

Primary Contributions

This work describes 3 primary contributions. First, we created caBERT: a BERT-based language model for comprehension of cancer pathology reports. We are aware of only 1 other attempt to create a pathology- and oncology-specific BERT language model [45]. That study included 290,438 pathology reports created between 2005 and 2015 from a tertiary teaching hospital in the United States. However, only 8870 of these reports were cases involving cancer. Our study included 275,605 pathology reports from patients with cancer diagnosed or treated at Moffitt. The larger corpus of cancer-specific reports should help our system achieve higher performance levels with cancer-related NLP tasks. However, because we did not have access to the system described in the study by Ma et al [45], a direct comparison with ours was not possible.

Second, we created a new Q&A system to extract tumor site and histology descriptions from free-text pathology reports. This is the first functional Q&A system for extracting

information from pathology reports that we are aware of. The Q&A format has 2 important benefits. First, it provides a user-friendly interface to the information extraction system. Second, incorporation of additional questions into our system is straightforward. With appropriate ground truth-labeled training data, this should allow us to extract additional data fields from free-text pathology reports.

Third, we created a new caBERT network, caBERTnet, to predict fine-grained ICD-O-3 site and histology codes using the answers extracted through the initial Q&A component. There has been considerable prior work using NLP methods to predict ICD-O codes from pathology reports [11,12,16,67]. Here, we compare our results to 5 of the most highly cited recent publications in this area.

Comparisons With Prior Work

Much of the prior work has focused on a single anatomical site or histology. For example, Coden et al [16] described a system to extract information on tumor site, histology, grade, lymph nodes, tumor size, and reporting date from free-text pathology reports of colon cancer. They achieved precision and recall

values of between 0.95 and 0.98 for both site and histology ICD-O codes. Their system used a rule-based NLP pipeline, with a large number of controlling parameters that required extensive manual tuning to obtain optimal results. In contrast, BERT-based NLP systems can both discover and tune the steps of a traditional NLP pipeline automatically [68]. This has significant advantages in terms of reduced effort; in addition, it allows these systems to be quickly retuned for data sets from other institutions or different applications through transfer learning [69].

The BERT system from Ma et al [45], mentioned earlier, was used to extract information on 15 *primary cancer sites*, 6 *cancer reasons*, and 3 *metastatic disease* states. In all, 11 of their cancer sites corresponded to ICD-O-3 group-level site classifications (eg, breast, lung, or bronchus). The others were broader groupings (eg, colorectal, upper gastrointestinal, and head and neck). Of their cancer reasons, 4 corresponded to ICD-O-3 group-level histology classifications (eg, melanoma and soft tissue sarcoma). The remaining two cancer reasons were very broad groupings: carcinoma and blastoma. They achieved accuracies on the full test set of 96.7% and 98.5% for cancer site and cancer reason, respectively. However, they did not predict ICD-O-3 group or fine-grained codes.

Nguyen et al [70] developed a system to monitor Health Level Seven International electronic pathology reports from across the state of Queensland in Australia. Their system relied on business rules and symbolic reasoning using Systematized Nomenclature of Medicine codes. They tuned their system using 201 pathology reports and tested it on 220 unseen reports. They extracted 8 different cancer characteristics from these reports. These characteristics included ICD-O-3 site codes (both fine-grained, Cxxx, and group, Cxx) and histological type. Their data set included 66 sites and 94 histologies. They achieved F1 scores of 61.1%, 73.2% and 63.7% on fine-grained site codes, group site codes, and histology codes, respectively.

Alawad et al [67] developed a multistage system of deep convolutional neural networks to extract the primary site, histological grade, and laterality from pathology reports. They achieved an F1 score of 77.5% over 12 ICD-O-3 site codes.

Qiu et al [11] also developed a deep convolutional neural network to extract ICD-O-3 codes from breast and lung cancer pathology reports. Training was based on 942 pathology reports annotated by cancer registry experts. The data set included 7 breast sites and 5 lung sites. Of the 12 sites, 6 had at least 50 samples per code. The remaining 6 sites had 10 to 50 samples each. They evaluated their system using a 10-fold cross-validation. Their overall F1 score for predicting tumor sites across all 12 ICD-O-3 codes was 72.2%.

Our study included a greater diversity of cancer cases than previous studies (214 site codes and 193 histology codes), while obtaining similar or better accuracy scores. Many of the site and histology codes in our training data set included ≤ 5 samples, whereas prior studies reported ≥ 10 training samples per code. Culling codes from the test set with ≤ 5 samples in the training set reduced the size of our test data set by 6.39% (130/2034; 130 pathology reports). However, this increased our top-1 accuracy on the test data to 73.84% (1406/1904; +2.26%) and

85.29% (1624/1904; +1.42%) for site and histology, respectively. Our system also ranks and reports the top 5 predictions for ICD-O-3 site and histology codes. This has useful clinical applications: often there is a degree of uncertainty or *hedging* in pathology reports [16]. Listing the top 5 predicted codes could help to reduce this uncertainty. For example, an artificial intelligence–assisted abstraction system that provides the top 5 predicted ICD-O-3 codes for a particular pathology report (in a pull-down menu, for example) could aid the process of abstraction and enhance the workflow in cancer registries. Our top 5 accuracies for fine-grained codes with ≥ 5 training samples were 92.95% (1794/1930) and 96.01% (1853/1930) for site and histology, respectively.

Additional Insights From the Results

Figure 5 shows the top 5 results at various levels of rare-code elimination from the test data set, and it provides 3 additional insights. First, as N increased from 1 to 5, the improvement in accuracy for sites was larger than that for histologies. This suggests that there is more uncertainty predicting site codes than in predicting histology codes. Second, eliminating rare codes, for example, going from E=0 (green lines) to E=5 (blue lines; Figure 5), improved site accuracy more than it improved histology accuracy. This suggests that site prediction was more dependent on sample size. Third, site accuracy failed to improve for E>20. This suggests that 20 samples per code were required to maximize site code prediction accuracy.

The overall accuracy for predicting site group codes was 93.53% (1895/2026) (Figure 6). Nevertheless, several site group codes had accuracies <80%. Here, we will discuss the *Other* group (52/90, 58%, accuracy), along with two of the site group codes with the lowest accuracies: *C49 Connective, Subcutaneous, and Other Soft Tissues* (15/29, 52%, accuracy) and *C16 Stomach* (14/27, 52%, accuracy).

The *Other* site category included 27 group codes. Together, these group codes contained 61 fine-grained codes with at least one sample pathology report each in the training data set, as determined by the MCR. The mean and median number of reports in the training data set for each fine-grained code in the *Other* category were 6.3 (SD 6.3) and 4 (IQR 5), respectively. Consequently, caBERTnet accuracy on these rare sites was likely limited by the availability of training data.

caBERTnet failed to predict the MCR site code for 14 test cases in the group *C49 Connective, Subcutaneous, and Other Soft Tissues*. We manually reviewed 50% (7/14) of these cases, all of which were labeled by the MCR as soft tissue of the limb, shoulder, and hip or pelvis (codes *C491*, *C492*, and *C495*). In 14% (1/7) of these cases, the information required to determine the correct site was not present in the pathology report text. In these situations, CTRs would use additional information in the patient record. However, this information was not available to caBERTnet. In the remaining 86% (6/7) of the cases, the pathology report described characteristics of a lesion that had metastasized from the limb, shoulder, hip, or pelvis to another location. The MCR recorded the originating organ as the tumor site, whereas caBERTnet predicted the metastasis site.

caBERTnet failed to predict the MCR site code for 13 test cases in the group *C16 Stomach*. All these cases were labeled by the MCR as *C160 Cardia, NOS*, and by caBERTnet as lesions of the lower third of the esophagus (*C155*; 12/13, 92%) or as overlapping lesions of the esophagus (*C158*; 1/13, 8%). The MCR labels are due to a rule in the American Joint Committee on Cancer Staging Manual, Eighth Edition [71]. On page 189 in that manual it states as follows:

Cancers involving the Esophagogastric Junction (EGJ) that have their epicenter within the proximal 2 cm of the cardia (Siewert types I/II) are to be staged as esophageal cancers. Cancers whose epicenter is more than 2 cm distal from the EGJ, even if the EGJ is involved, will be staged using the stomach cancer TNM (primary tumor, lymph nodes, and distant metastases) and stage groupings (see Chapter 17).

The pathology reports on these cases did not mention the spatial location of the tumor sample in relation to the EGJ. Consequently, measurement of the tumor location in pretreatment imaging was required to determine the correct tumor site code.

The overall accuracy for predicting histology group codes was 97.7% (Figure 7). Only one group code had an accuracy <80%: *801-804 Epithelial Neoplasms, NOS* (77.8%). caBERTnet failed to predict the MCR histology code for 5 of these cases. In 80% (4/5) of these cases, the pathology report was based on histology at the metastatic site of disease. The MCR coded these as the histology of the originating tumor, whereas caBERTnet predicted the histology at the metastatic site. In 20% (1/5) of the cases, the information required to determine the correct histology code was not present in the pathology report and required the CTR to conduct a review of the patient medical record.

The last case was quite interesting because the pathology report included an initial intraoperative diagnosis that disagreed with the final diagnosis. The former indicated a histological type of *spindle cell carcinoma*. The latter included the following statements: “the differential diagnosis includes sarcomatoid carcinoma and inflammatory myofibroblastic tumor...the histomorphologic and immunoprofile support the diagnosis of sarcomatoid carcinoma.” The MCR coded the histology as *8032/3 spindle cell carcinoma, NOS*, based on the intraoperative statements, whereas caBERTnet predicted *8033/3 pseudosarcomatous carcinoma*. The phrase *sarcomatoid carcinoma* is an alternative form of the ICD-O-3 preferred phrase *pseudosarcomatous carcinoma*. Although caBERTnet’s prediction did not agree with that of the MCR, downstream applications may still value automatic prediction and codification of the final diagnosis.

Potential Applications

There are multiple potential applications of caBERTnet at Moffitt. For example, there is a delay of several months between initial pathology report dictation and CTR abstraction because the CTRs typically wait for enough time to have elapsed for the first course treatment to have been administered to minimize the number of times they have to review the medical record.

caBERTnet can be used to extract information from pathology reports in a timelier way, thus facilitating the use of the data for clinical pathway reporting and screening for clinical trials. Furthermore, CTRs only abstract the subset of pathology reports associated with the cancer diagnosis and first-course treatment. caBERTnet could be used to extract information from pathology reports associated with subsequent biopsies and surgeries that would never be manually curated by the CTRs. To facilitate these use cases, we plan to extract tumor site and histology information close to real time and link these values to other patient data stored in our analytics platform. These data can be incorporated into real-time dashboards and data sets for a wide range of decision support and research applications.

We do not believe that caBERTnet will replace CTRs at cancer clinics. Many complex, difficult, and rare cases require intuition and information outside of the pathology report to determine the correct coding. These cases are beyond the scope of an NLP tool. However, caBERTnet may help simplify and accelerate MCR workflows. For example, caBERTnet could preprocess pathology reports to identify the top 5 site and histology ICD-O-3 codes and their corresponding phrases. The phrases could then be highlighted within the report body. In addition, two pull-down menus could be prepopulated with top 5 code predictions: one for site and the other for histology. The CTR could then quickly choose a code from either pull-down menu. If the correct code was not among the top 5, then the CTR would resort to their current workflow, entering this information by hand.

Limitations of Approach

Although there are immediate applications of the caBERTnet model within our internal workflows, there are a number of aspects of our modeling approach that limit the application of caBERTnet to other use cases. In this section, we provide an overview of several important limitations that we believe should be considered before caBERTnet implementation.

The most critical limitations correspond to issues related to our data curation and preprocessing approaches. In the curation of our Moffitt fine-tuning data set, we restricted the available reports to only those containing a single primary tumor diagnosis. Although this was partly imposed by the nature of the SQuAD Q&A task (which expects a single answer to each question asked of each input), it nonetheless limits the generalizability of our model to reports containing multiple (or zero) positive diagnoses. We are exploring methods of mitigating these issues within the current setup by adjusting the likelihood thresholds for output predictions that could be used to screen out reports with no diagnosis. Another limitation inherent to a Q&A system is the necessity of knowing the precise span of text corresponding to the answer to each question asked of each report. Owing to the sheer number of reports in our fine-tuning data set, it was infeasible to manually curate answer labels. To circumvent this issue, we chose to create our own automated system to determine the answer text in each report. Any such automated preprocessing necessarily leaves a fingerprint on downstream tasks. The drawbacks of our approach in particular relate to the restriction of answers from a predetermined list of possibilities for each site and histology.

Although these phrase sets were diverse (and our approach even allowed for permutations of phrasing within these terms), this process nevertheless limits the allowed terminology and is necessarily incomplete.

In addition, a single pathology report may reference several tissue samples (which can be from related or distant sites from the actual site of diagnosis). Although we limited our automated answer-search preprocessing algorithm to only find terms associated with the known diagnosis label for each report, it is possible that the answer found in the report text by the preprocessing corresponds to a false answer from a different sample in the report; this is particularly relevant for cases with multiple samples from related sites (eg, upper outer quadrant breast vs upper inner quadrant breast). When the 2 samples are in the same site (or histology) group, this issue is avoided by outputting the group code; however, this does not help when the 2 samples belong to different groups altogether.

Future Directions

We plan to continue development of caBERTnet. Of particular interest is extending the system with additional questions and MCR-derived ground truth labels to train it to extract additional tumor characteristics. These include grade, size, involvement of lymph nodes, primary or metastatic status, presence or absence of molecular markers, and others. caBERTnet could also be customized to extract information on hematological malignancies.

A caBERTnet-assisted MCR abstraction tool could also be used for active [72] or human-in-the-loop [73] learning. Briefly, this approach uses human-labeled data to improve the performance of machine learning algorithms over time. It is particularly useful when the subject matter expert (a CTR in our case) provides labels for cases with low-confidence predictions by the machine learning algorithm. However, it would require careful engineering to avoid common pitfalls and ensure seamless operation [74].

The system's accuracy on rare sites and histologies could be improved with additional training data. A potential option may

be to collaborate with other academic cancer centers; distributing caBERTnet for training on local pathology reports at other sites based on ground truth labeling of training and test data sets from highly standardized registry data or the application of federated multitask learning [75] that distributes copies of a central model to multiple spoke sites for tuning of the central model on local data could allow vast improvements in caBERTnet accuracy. Using these methods, information learned at local sites (eg, model weights) is transmitted back to the central node where the information is combined in a pluralistic way that avoids the need to impose consensus on the data distributions at the spoke sites. This allows for both heterogeneity in local data and broad generalizability of the central model. That said, the most effective way to protect private health information when sharing such models remains an unsolved problem, and these kinds of expansions would depend on the development of validated privacy schemes specific to BERT.

Conclusions

Our new NLP system, caBERTnet, is built around a network of 3 cooperating BERT instances. On a sequestered test data set, it produced top 5 accuracies of 92.95% (1794/1930) and 96.01% (1853/1930) for fine-grained ICD-O-3 site and histology codes, respectively. This level of accuracy is on par with existing systems in the literature, while also being accurate over a broader range of site and histology groups.

Pathology report abstraction systems such as caBERTnet cannot be expected to achieve performance on par with CTRs who abstract data ultimately incorporated into population-based cancer registries [76], given the vast amount of ancillary data from within the electronic health record that is required for thorough abstraction and the subtle nuances associated with coding guidelines. However, caBERTnet could expedite access to timely pathology data needed for disease surveillance, cohort identification, and clinical trial matching. Furthermore, it can improve existing workflows, serving as a valuable step toward the ultimate goal of a mostly automated abstraction system.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Additional information on the CancerBERT and CancerBERT network model training and performance as well as an elaboration on the acceptable phrases used to curate labels for pathology reports in the question-and-answer stage.

[\[DOCX File , 360 KB - jmir_v24i3e27210_app1.docx \]](#)

References

1. Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013 Apr 03;309(13):1351-1352. [doi: [10.1001/jama.2013.393](https://doi.org/10.1001/jama.2013.393)] [Medline: [23549579](https://pubmed.ncbi.nlm.nih.gov/23549579/)]
2. Pratt A, Thomas L. An information processing system for pathology data. *Pathol Annul* 1966;1.
3. Dunham GS, Pacak MG, Pratt AW. Automatic indexing of pathology data. *J Am Soc Inf Sci* 1978 Mar;29(2):81-90. [doi: [10.1002/asi.4630290207](https://doi.org/10.1002/asi.4630290207)] [Medline: [10318395](https://pubmed.ncbi.nlm.nih.gov/10318395/)]
4. Meystre S, Haug PJ. Natural language processing to extract medical problems from electronic clinical documents: performance evaluation. *J Biomed Inform* 2006 Dec;39(6):589-599 [FREE Full text] [doi: [10.1016/j.jbi.2005.11.004](https://doi.org/10.1016/j.jbi.2005.11.004)] [Medline: [16359928](https://pubmed.ncbi.nlm.nih.gov/16359928/)]

5. Murff HJ, FitzHenry F, Matheny ME, Gentry N, Kotter KL, Crimin K, et al. Automated identification of postoperative complications within an electronic medical record using natural language processing. *JAMA* 2011 Aug 24;306(8):848-855. [doi: [10.1001/jama.2011.1204](https://doi.org/10.1001/jama.2011.1204)] [Medline: [21862746](https://pubmed.ncbi.nlm.nih.gov/21862746/)]
6. Wu S, Roberts K, Datta S, Du J, Ji Z, Si Y, et al. Deep learning in clinical natural language processing: a methodical review. *J Am Med Inform Assoc* 2020 Mar 01;27(3):457-470 [FREE Full text] [doi: [10.1093/jamia/ocz200](https://doi.org/10.1093/jamia/ocz200)] [Medline: [31794016](https://pubmed.ncbi.nlm.nih.gov/31794016/)]
7. Datta S, Bernstam EV, Roberts K. A frame semantic overview of NLP-based information extraction for cancer-related EHR notes. *J Biomed Inform* 2019 Dec;100:103301 [FREE Full text] [doi: [10.1016/j.jbi.2019.103301](https://doi.org/10.1016/j.jbi.2019.103301)] [Medline: [31589927](https://pubmed.ncbi.nlm.nih.gov/31589927/)]
8. Napolitano C, Fox C, Middleton R, Connolly D. Pattern-based information extraction from pathology reports for cancer registration. *Cancer Causes Control* 2010 Nov;21(11):1887-1894. [doi: [10.1007/s10552-010-9616-4](https://doi.org/10.1007/s10552-010-9616-4)] [Medline: [20652738](https://pubmed.ncbi.nlm.nih.gov/20652738/)]
9. Soysal J, Warner JL, Denny JC, Xu H. Identifying metastases-related information from pathology reports of lung cancer patients. *AMIA Jt Summits Transl Sci Proc* 2017;2017:268-277 [FREE Full text] [Medline: [28815141](https://pubmed.ncbi.nlm.nih.gov/28815141/)]
10. Lee H, Song HJ, Yoon E, Park SB, Park SH, Seo JW, et al. Automated extraction of Biomarker information from pathology reports. *BMC Med Inform Decis Mak* 2018 May 21;18(1):29 [FREE Full text] [doi: [10.1186/s12911-018-0609-7](https://doi.org/10.1186/s12911-018-0609-7)] [Medline: [29783980](https://pubmed.ncbi.nlm.nih.gov/29783980/)]
11. Qiu J, Yoon H, Fearn P, Tourassi G. Deep learning for automated extraction of primary sites from cancer pathology reports. *IEEE J Biomed Health Inform* 2018 Jan;22(1):244-251. [doi: [10.1109/JBHI.2017.2700722](https://doi.org/10.1109/JBHI.2017.2700722)] [Medline: [28475069](https://pubmed.ncbi.nlm.nih.gov/28475069/)]
12. Gao M, Young MT, Qiu JX, Yoon HJ, Christian JB, Fearn PA, et al. Hierarchical attention networks for information extraction from cancer pathology reports. *J Am Med Inform Assoc* 2018 Mar 01;25(3):321-330 [FREE Full text] [doi: [10.1093/jamia/ocx131](https://doi.org/10.1093/jamia/ocx131)] [Medline: [29155996](https://pubmed.ncbi.nlm.nih.gov/29155996/)]
13. Pattison D, Denhalter D, Patterson O, DuVall S, Seigne J, Sirovich B, et al. MP31-14 leveraging bladder cancer pathology reports for research: gleaning meaning despite widely variable language. *J Urol* 2016 Apr;195(4S):e424-e425 [FREE Full text] [doi: [10.1016/j.juro.2016.02.1267](https://doi.org/10.1016/j.juro.2016.02.1267)]
14. Buckley JM, Coopey SB, Sharko J, Polubriaginof F, Drohan B, Belli AK, et al. The feasibility of using natural language processing to extract clinical information from breast pathology reports. *J Pathol Inform* 2012;3:23 [FREE Full text] [doi: [10.4103/2153-3539.97788](https://doi.org/10.4103/2153-3539.97788)] [Medline: [22934236](https://pubmed.ncbi.nlm.nih.gov/22934236/)]
15. Tang L, Ouyang L, Li C, He Y, Griffin M, Taghian A, et al. Machine learning to parse breast pathology reports in Chinese. *Breast Cancer Res Treat* 2018 Jun;169(2):243-250. [doi: [10.1007/s10549-018-4668-3](https://doi.org/10.1007/s10549-018-4668-3)] [Medline: [29380208](https://pubmed.ncbi.nlm.nih.gov/29380208/)]
16. Coden G, Savova G, Sominsky I, Tanenblatt M, Masanz J, Schuler K, et al. Automatically extracting cancer disease characteristics from pathology reports into a Disease Knowledge Representation Model. *J Biomed Inform* 2009 Oct;42(5):937-949 [FREE Full text] [doi: [10.1016/j.jbi.2008.12.005](https://doi.org/10.1016/j.jbi.2008.12.005)] [Medline: [19135551](https://pubmed.ncbi.nlm.nih.gov/19135551/)]
17. Nguyen M. Structured pathology reporting for cancer from free text: lung cancer case study. *Electron J Health Informatics* 2012;7(1) [FREE Full text]
18. Yoon HJ, Robinson S, Christian JB, Qiu JX, Tourassi GD. Filter pruning of Convolutional Neural Networks for text classification: a case study of cancer pathology report comprehension. In: Proceedings of the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). 2018 Presented at: 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI); Mar 4-7, 2018; Las Vegas, NV, USA. [doi: [10.1109/bhi.2018.8333439](https://doi.org/10.1109/bhi.2018.8333439)]
19. Schroeck O, Patterson O, Alba P, DuVall S, Sirovich B, Robertson D, et al. MP32-12 harnessing full text pathology data from the electronic health record to advance bladder cancer care – development of a natural language processing system to generate longitudinal pathology data. *J Urol* 2017 Apr;197(4S):e413 [FREE Full text] [doi: [10.1016/j.juro.2017.02.987](https://doi.org/10.1016/j.juro.2017.02.987)]
20. Schadow G, McDonald CJ. Extracting structured information from free text pathology reports. *AMIA Annu Symp Proc* 2003:584-588 [FREE Full text] [Medline: [14728240](https://pubmed.ncbi.nlm.nih.gov/14728240/)]
21. Burger G, Abu-Hanna A, de Keizer N, Cornet R. Natural language processing in pathology: a scoping review. *J Clin Pathol* 2016 Jul 22;jclinpath-2016-203872. [doi: [10.1136/jclinpath-2016-203872](https://doi.org/10.1136/jclinpath-2016-203872)] [Medline: [27451435](https://pubmed.ncbi.nlm.nih.gov/27451435/)]
22. Névéal A, Zweigenbaum P, Section Editors for the IMIA Yearbook Section on Clinical Natural Language Processing. Expanding the diversity of texts and applications: findings from the section on clinical natural language processing of the international medical informatics association yearbook. *Yearb Med Inform* 2018 Aug;27(1):193-198 [FREE Full text] [doi: [10.1055/s-0038-1667080](https://doi.org/10.1055/s-0038-1667080)] [Medline: [30157523](https://pubmed.ncbi.nlm.nih.gov/30157523/)]
23. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of NAACL-HLT 2019. 2019 Presented at: Proceedings of NAACL-HLT 2019; Jun 2 - 7, 2019; Minneapolis, Minnesota URL: <https://aclanthology.org/N19-1423.pdf>
24. Wang A, Singh A, Michael J, Hill F, Levy O, Bowman S. GLUE: a multi-task benchmark and analysis platform for natural language understanding. In: Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP. 2018 Presented at: Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP; Nov 1, 2018; Brussels, Belgium. [doi: [10.18653/v1/w18-5446](https://doi.org/10.18653/v1/w18-5446)]
25. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training. *Glue Benchmark*. URL: <https://www.cs.ubc.ca/~amuham01/LING530/papers/radford2018improving.pdf> [accessed 2022-02-12]
26. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. *OpenAI Blog*. URL: https://www.ceid.upatras.gr/webpages/faculty/zaro/teaching/alg-ds/PRESENTATIONS/PAPERS/2019-Radford-et-al_Language-Models-Are-Unsupervised-Multitask-%20Learners.pdf [accessed 2022-02-12]

27. Dai Z, Yang Z, Yang Y, Carbonell J, Le Q, Salakhutdinov R. Transformer-XL: attentive language models beyond a fixed-length context. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019 Presented at: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; Jul 28 - Aug 2, 2019; Florence, Italy URL: <https://aclanthology.org/P19-1285.pdf> [doi: [10.18653/v1/p19-1285](https://doi.org/10.18653/v1/p19-1285)]
28. Yang Z, Dai Z, Yang Y, Carbonell J, Salakhutdinov R, Le Q. XLNet: generalized autoregressive pretraining for language understanding. ArXiv. Preprint posted online January 2, 2020. [[FREE Full text](#)]
29. Lan Z, Chen M, Goodman S, Gimpel K, Sharma P, Soricut R. ALBERT: a Lite BERT for self-supervised learning of language representations. ArXiv. Preprint posted online February 9, 2020. [[FREE Full text](#)]
30. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D. RoBERTa: a robustly optimized BERT pretraining approach. ArXiv. Preprint posted online July 26, 2019. [[FREE Full text](#)]
31. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv. Preprint posted online March 1, 2020. [[FREE Full text](#)]
32. Lample G, Conneau A. Cross-lingual language model pretraining. ArXiv. Preprint posted online January 22, 2019. [[FREE Full text](#)] [doi: [10.1090/mbk/121/79](https://doi.org/10.1090/mbk/121/79)]
33. Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M. Exploring the limits of transfer learning with a unified text-to-text transformer. ArXiv. Preprint posted online July 28, 2020. [[FREE Full text](#)]
34. Ju Y, Zhao F, Chen S, Zheng B, Yang X, Liu Y. Technical report on conversational question answering. ArXiv Preprint posted online September 29, 2019. [[FREE Full text](#)]
35. Han W, Zhang Z, Zhang Y, Yu J, Chiu C, Qin J. ContextNet: improving convolutional neural networks for automatic speech recognition with global context. In: Proceedings of the INTERSPEECH 2020. 2020 Presented at: INTERSPEECH 2020; Oct 25-29, 2020; Shanghai, China URL: <http://arxiv.org/abs/2005.03191> [doi: [10.21437/interspeech.2020-2059](https://doi.org/10.21437/interspeech.2020-2059)]
36. Peng X, Long G, Shen T, Wang S, Jiang J, Zhang C. BiteNet: bidirectional temporal encoder network to predict medical outcomes. In: Proceedings of the 2020 IEEE International Conference on Data Mining (ICDM). 2020 Presented at: 2020 IEEE International Conference on Data Mining (ICDM); Nov 17- 20, 2020; Sorrento, Italy. [doi: [10.1109/icdm50108.2020.00050](https://doi.org/10.1109/icdm50108.2020.00050)]
37. Li F, Jin Y, Liu W, Rawat BP, 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](https://doi.org/10.2196/14830)] [Medline: [31516126](https://pubmed.ncbi.nlm.nih.gov/31516126/)]
38. Huang K, Altsaar J, Ranganath R. ClinicalBERT: modeling clinical notes and predicting hospital readmission. ArXiv. Preprint posted online November 29, 2020. [[FREE Full text](#)]
39. Tahayori B, Chini-Foroush N, Akhlaghi H. Advanced natural language processing technique to predict patient disposition based on emergency triage notes. *Emerg Med Australas* 2020 Oct 11 (forthcoming). [doi: [10.1111/1742-6723.13656](https://doi.org/10.1111/1742-6723.13656)] [Medline: [33043570](https://pubmed.ncbi.nlm.nih.gov/33043570/)]
40. Blinov P, Avetisian M, Kokh V, Umerenkov D, Tuzhilin A. Predicting clinical diagnosis from patients electronic health records using BERT-based neural networks. In: *Artificial Intelligence in Medicine*. Cham: Springer; 2020.
41. Xu D, Gopale M, Zhang J, Brown K, Begoli E, Bethard S. Unified Medical Language System resources improve sieve-based generation and Bidirectional Encoder Representations from Transformers (BERT)-based ranking for concept normalization. *J Am Med Inform Assoc* 2020 Oct 01;27(10):1510-1519 [[FREE Full text](#)] [doi: [10.1093/jamia/ocaa080](https://doi.org/10.1093/jamia/ocaa080)] [Medline: [32719838](https://pubmed.ncbi.nlm.nih.gov/32719838/)]
42. Kandpal P, Jasnani K, Raut R, Bhorge S. Contextual chatbot for healthcare purposes (using Deep Learning). In: Proceedings of the 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4). 2020 Presented at: 2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4); Jul 27-28, 2020; London, UK. [doi: [10.1109/worlds450073.2020.9210351](https://doi.org/10.1109/worlds450073.2020.9210351)]
43. Zhang L, Fan H, Peng C, Rao G, Cong Q. Sentiment analysis methods for HPV vaccines related tweets based on transfer learning. *Healthcare (Basel)* 2020 Aug 28;8(3):307 [[FREE Full text](#)] [doi: [10.3390/healthcare8030307](https://doi.org/10.3390/healthcare8030307)] [Medline: [32872330](https://pubmed.ncbi.nlm.nih.gov/32872330/)]
44. Shang J, Ma T, Xiao C, Sun J. Pre-training of graph augmented transformers for medication recommendation. ArXiv. Preprint posted online November 27, 2019. [[FREE Full text](#)] [doi: [10.24963/ijcai.2019/825](https://doi.org/10.24963/ijcai.2019/825)]
45. Ma R, Chen P, Li G, Weng W, Lin A, Gadepalli K. Human-centric metric for accelerating pathology reports annotation. ArXiv. Preprint posted online November 12, 2019. [[FREE Full text](#)]
46. Grishman R, Hirschman L. Question answering from natural language medical data bases. *Artificial Intelligence* 1978 Aug;11(1-2):25-43. [doi: [10.1016/0004-3702\(78\)90011-5](https://doi.org/10.1016/0004-3702(78)90011-5)]
47. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, et al. PyTorch: an imperative style, high-performance deep learning library. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019 Presented at: Proceedings of the 33rd International Conference on Neural Information Processing Systems; Dec 8-14, 2019; Vancouver, Canada. [doi: [10.5555/3454287.3455008](https://doi.org/10.5555/3454287.3455008)]
48. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. 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: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations; Nov 16-20, 2020; -. [doi: [10.18653/v1/2020.emnlp-demos.6](https://doi.org/10.18653/v1/2020.emnlp-demos.6)]

49. Jain A, Meenachi N, Venkatraman B. NukeBERT: a pre-trained language model for low resource nuclear domain. ArXiv. Preprint posted online August 19, 2020. [FREE Full text]
50. Neumann M, King D, Beltagy I, Ammar W. ScispaCy: fast and robust models for biomedical natural language processing. In: Proceedings of the 18th BioNLP Workshop and Shared Task. 2019 Presented at: Proceedings of the 18th BioNLP Workshop and Shared Task; Aug 1, 2019; Florence, Italy. [doi: [10.18653/v1/w19-5034](https://doi.org/10.18653/v1/w19-5034)]
51. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A. Attention is all you need. In: Proceedings of the 31st Conference on Neural Information Processing Systems. 2017 Presented at: 31st Conference on Neural Information Processing Systems; Dec 4 - 9, 2017; Long Beach, CA, USA URL: <https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf>
52. Clark K, Khandelwal U, Levy O, Manning C. What does BERT look at? An Analysis of BERT's Attention. In: Proceedings of the Second BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP. 2019 Presented at: Proceedings of the Second BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP; Aug 1, 2019; Florence, Italy. [doi: [10.18653/v1/w19-4828](https://doi.org/10.18653/v1/w19-4828)]
53. Zhang Z, Sabuncu M. Generalized cross entropy loss for training deep neural networks with noisy labels. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems. 2018 Presented at: Proceedings of the 32nd International Conference on Neural Information Processing Systems; Dec 3 - 8, 2018; Montréal Canada. [doi: [10.5555/3327546.3327555](https://doi.org/10.5555/3327546.3327555)]
54. Lee J, Yoon W, Kim S, Kim D, Kim S, So C, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 2020 Feb 15;36(4):1234-1240 [FREE Full text] [doi: [10.1093/bioinformatics/btz682](https://doi.org/10.1093/bioinformatics/btz682)] [Medline: [31501885](https://pubmed.ncbi.nlm.nih.gov/31501885/)]
55. Alsentzer E, Murphy J, Boag W, Weng W, Jin D, Naumann T. Publicly available clinical BERT embeddings. In: Proceedings of the 2nd Clinical Natural Language Processing Workshop. 2019 Presented at: 2nd Clinical Natural Language Processing Workshop; Jun 7, 2019; Minneapolis, USA URL: <http://arxiv.org/abs/1904.03323> [doi: [10.18653/v1/w19-1909](https://doi.org/10.18653/v1/w19-1909)]
56. Johnson AE, 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](https://doi.org/10.1038/sdata.2016.35)] [Medline: [27219127](https://pubmed.ncbi.nlm.nih.gov/27219127/)]
57. Rajpurkar P, Zhang J, Lopyrev K, Liang P. SQuAD: 100,000+ questions for machine comprehension of text. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2016 Presented at: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP); Nov 1–5, 2016; Austin, Texas, USA URL: <http://arxiv.org/abs/1606.05250> [doi: [10.18653/v1/d16-1264](https://doi.org/10.18653/v1/d16-1264)]
58. Tsatsaronis G, Balikas G, Malakasiotis P, Partalas I, Zschunke M, Alvers MR, et al. An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition. *BMC Bioinformatics* 2015 Apr 30;16:138 [FREE Full text] [doi: [10.1186/s12859-015-0564-6](https://doi.org/10.1186/s12859-015-0564-6)] [Medline: [25925131](https://pubmed.ncbi.nlm.nih.gov/25925131/)]
59. Yoon W, Lee J, Kim D, Jeong M, Kang J. Pre-trained language model for biomedical question answering. ArXiv. Preprint posted online September 18, 2019. [FREE Full text] [doi: [10.1007/978-3-030-43887-6_64](https://doi.org/10.1007/978-3-030-43887-6_64)]
60. Hawhee V, Jones J, Stewart S, Lucia K, Rollison DE. Quality assurance and continuing education: a cyclic approach for maintaining high quality data in a high volume cancer registry. *Cancer Control* 2020;27(3):1073274820946794 [FREE Full text] [doi: [10.1177/1073274820946794](https://doi.org/10.1177/1073274820946794)] [Medline: [32755235](https://pubmed.ncbi.nlm.nih.gov/32755235/)]
61. International Classification of Diseases for Oncology. Geneva: World Health Organization; 2013.
62. ICD-O-3 Coding Materials. NIH National Cancer institute. URL: <https://seer.cancer.gov/icd-o-3/> [accessed 2022-02-12]
63. Site-specific modules. NIH National Cancer Institute. URL: https://training.seer.cancer.gov/modules_site_spec.html [accessed 2022-02-12]
64. International classification of diseases for oncology (ICD-O). International Agency for Research on Cancer. URL: http://www.iacr.com.fr/index.php?option=com_content&view=category&layout=blog&id=100&Itemid=577 [accessed 2022-02-12]
65. ICD-O-3.1-NCIt Mapping Files. National Institutes of Health. URL: https://evs.nci.nih.gov/ftp1/NCI_Thesaurus/Mappings/ICD-O-3_Mappings/About.html [accessed 2022-02-12]
66. Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Commun ACM* 2017 May 24;60(6):84-90. [doi: [10.1145/3065386](https://doi.org/10.1145/3065386)]
67. Alawad M, Yoon H, Tourassi G. Coarse-to-fine multi-task training of convolutional neural networks for automated information extraction from cancer pathology reports. In: Proceedings of the 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI). 2018 Presented at: IEEE EMBS International Conference on Biomedical Health Informatics (BHI). ieeexplore.ieee.org; . pp. 218?221; Mar 4-7, 2018; Las Vegas, NV, USA. [doi: [10.1109/bhi.2018.8333408](https://doi.org/10.1109/bhi.2018.8333408)]
68. Tenney I, Das D, Pavlick E. BERT rediscovers the classical NLP pipeline. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (2019). 2019 Presented at: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (2019); Jul 28-Aug 2, 2019; Florence, Italy URL: <http://arxiv.org/abs/1905.05950> [doi: [10.18653/v1/p19-1452](https://doi.org/10.18653/v1/p19-1452)]
69. Pan SJ, Yang Q. A survey on transfer learning. *IEEE Trans Knowl Data Eng* 2010 Oct;22(10):1345-1359. [doi: [10.1109/TKDE.2009.191](https://doi.org/10.1109/TKDE.2009.191)]
70. Nguyen AN, Moore J, O'Dwyer J, Philpot S. Assessing the utility of automatic cancer registry notifications data extraction from free-text pathology reports. *AMIA Annu Symp Proc* 2015;2015:953-962 [FREE Full text] [Medline: [26958232](https://pubmed.ncbi.nlm.nih.gov/26958232/)]

71. Amin MB, Edge SB. AJCC Cancer Staging Manual. Cham: Springer; 2017.
72. Active learning literature survey. University of Wisconsin-Madison Department of Computer Sciences. 2009. URL: <https://minds.wisconsin.edu/handle/1793/60660> [accessed 2022-02-12]
73. Holzinger A. Interactive machine learning for health informatics: when do we need the human-in-the-loop? *Brain Inform* 2016 Jun;3(2):119-131 [FREE Full text] [doi: [10.1007/s40708-016-0042-6](https://doi.org/10.1007/s40708-016-0042-6)] [Medline: [27747607](https://pubmed.ncbi.nlm.nih.gov/27747607/)]
74. Settles B. From theories to queries: active learning in practice. *JMLR Workshop Conference Proc* 2011;16:1-18 [FREE Full text]
75. Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated learning for healthcare informatics. *J Healthc Inform Res* 2020 Nov 12:1-19 [FREE Full text] [doi: [10.1007/s41666-020-00082-4](https://doi.org/10.1007/s41666-020-00082-4)] [Medline: [33204939](https://pubmed.ncbi.nlm.nih.gov/33204939/)]
76. Thoburn KK, German RR, Lewis M, Nichols P, Ahmed F, Jackson-Thompson J. Case completeness and data accuracy in the Centers for Disease Control and Prevention's National Program of Cancer Registries. *Cancer* 2007 Apr 15;109(8):1607-1616 [FREE Full text] [doi: [10.1002/cncr.22566](https://doi.org/10.1002/cncr.22566)] [Medline: [17343277](https://pubmed.ncbi.nlm.nih.gov/17343277/)]

Abbreviations

BERT: bidirectional encoder representations from transformers
BioASQ: Biomedical Semantic Indexing and Question Answering
caBERTnet: CancerBERT network
CTR: certified tumor registrar
EGJ: esophagogastric junction
EMR: electronic medical record
ICD-O-3: International Classification of Diseases for Oncology, Third Edition
MCR: Moffitt Cancer Registry
MIMIC-III: Medical Information Mart for Intensive Care, version 3
NLP: natural language processing
NOS: not otherwise specified
Q&A: question and answer
SEER: Surveillance, Epidemiology, and End Results
SQuAD: Stanford Question Answering Dataset

Edited by R Kukafka; submitted 16.01.21; peer-reviewed by K Roberts, R De Carvalho, A Louren; comments to author 30.07.21; revised version received 22.10.21; accepted 10.11.21; published 23.03.22.

Please cite as:

Mitchell JR, Szepietowski P, Howard R, Reisman P, Jones JD, Lewis P, Fridley BL, Rollison DE
A Question-and-Answer System to Extract Data From Free-Text Oncological Pathology Reports (CancerBERT Network): Development Study
J Med Internet Res 2022;24(3):e27210
URL: <https://www.jmir.org/2022/3/e27210>
doi: [10.2196/27210](https://doi.org/10.2196/27210)
PMID: [35319481](https://pubmed.ncbi.nlm.nih.gov/35319481/)

©Joseph Ross Mitchell, Phillip Szepietowski, Rachel Howard, Phillip Reisman, Jennie D Jones, Patricia Lewis, Brooke L Fridley, Dana E Rollison. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 23.03.2022. 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.

Original Paper

Privacy Preservation in Patient Information Exchange Systems Based on Blockchain: System Design Study

Sejong Lee^{1,2}, BS; Jaehyeon Kim^{2,3}, BS; Yongseok Kwon^{1,2}, BS; Teasung Kim¹, BS; Sunghyun Cho¹, PhD

¹Department of Computer Science and Engineering, Hanyang University, Ansan, Republic of Korea

²Major in Bio-Artificial Intelligence, Hanyang University, Ansan, Republic of Korea

³Department of Applied Artificial Intelligence, Hanyang University, Ansan, Republic of Korea

Corresponding Author:

Sunghyun Cho, PhD

Department of Computer Science and Engineering

Hanyang University

55, Hanyangdaehak-ro, Sangnok-gu, Ansan-si, Gyeonggi-do

Ansan, KS009

Republic of Korea

Phone: 82 31 400 5670

Email: chopro@hanyang.ac.kr

Abstract

Background: With the increasing sophistication of the medical industry, various advanced medical services such as medical artificial intelligence, telemedicine, and personalized health care services have emerged. The demand for medical data is also rapidly increasing today because advanced medical services use medical data such as user data and electronic medical records (EMRs) to provide services. As a result, health care institutions and medical practitioners are researching various mechanisms and tools to feed medical data into their systems seamlessly. However, medical data contain sensitive personal information of patients. Therefore, ensuring security while meeting the demand for medical data is a very important problem in the information age for which a solution is required.

Objective: Our goal is to design a blockchain-based decentralized patient information exchange (PIE) system that can safely and efficiently share EMRs. The proposed system preserves patients' privacy in the EMRs through a medical information exchange process that includes data encryption and access control.

Methods: We propose a blockchain-based EMR-sharing system that allows patients to manage their EMRs scattered across multiple hospitals and share them with other users. Our PIE system protects the patient's EMR from security threats such as counterfeiting and privacy attacks during data sharing. In addition, it provides scalability by using distributed data-sharing methods to quickly share an EMR, regardless of its size or type. We implemented simulation models using Hyperledger Fabric, an open source blockchain framework.

Results: We performed a simulation of the EMR-sharing process and compared it with previous works on blockchain-based medical systems to check the proposed system's performance. During the simulation, we found that it takes an average of 0.01014 (SD 0.0028) seconds to download 1 MB of EMR in our proposed PIE system. Moreover, it has been confirmed that data can be freely shared with other users regardless of the size or format of the data to be transmitted through the distributed data-sharing technique using the InterPlanetary File System. We conducted a security analysis to check whether the proposed security mechanism can effectively protect users of the EMR-sharing system from security threats such as data forgery or unauthorized access, and we found that the distributed ledger structure and re-encryption-based data encryption method can effectively protect users' EMRs from forgery and privacy leak threats and provide data integrity.

Conclusions: Blockchain is a distributed ledger technology that provides data integrity to enable patient-centered health information exchange and access control. PIE systems integrate and manage fragmented patient EMRs through blockchain and protect users from security threats during the data exchange process among users. To increase safety and efficiency in the EMR-sharing process, we used access control using security levels, data encryption based on re-encryption, and a distributed data-sharing scheme.

(*J Med Internet Res* 2022;24(3):e29108) doi:[10.2196/29108](https://doi.org/10.2196/29108)

KEYWORDS

electronic medical records; consortium blockchain; data security; medical data management; privacy preservation; smart contract; proxy re-encryption; patient-centered medical system; InterPlanetary File System

Introduction

Background

With the development of information and communication technology, the existing medical information system, which used paper charts to manage medical information such as patient treatment information and clinical results, changed to a digital-based medical information system. As of 2017, more than 94% of the hospitals in the United States have used digital health information systems [1,2]. The digital medical information system uses electronic medical records (EMRs) that store patient medical information (eg, patient demographics, progress notes, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports) in electronic document format for patient treatment and health management [3]. Moreover, health care practitioners use EMRs to provide improved health care to their patients through clinical decision support tools [4,5]. Of late, EMRs have been actively used in various fields (eg, medical artificial intelligence development [6-8], clinical trials [8-10], customized health care [11,12], and telemedicine [13,14]) by combining EMRs with the core technologies of the fourth industrial revolution. As EMRs are used widely, the value of, and demand for, EMRs continue to increase, and the size of the medical data market is also increasing every year [15,16]. There have been various attempts to share EMRs using networks to supply scarce EMRs, such as image-sharing networks [17,18] or health information exchange mechanisms [19-21]. However, the existing EMR-sharing systems centered on medical institutions that use a trusted third party (TTP) have security vulnerabilities and structural limitations. In the current EMR-sharing system, which manages data through a central database, overall service can be affected if a problem occurs in the database storing the data. Furthermore, if an attacker forges the EMR, it is difficult to determine whether the EMR has been forged if there are no original data for comparison, and in the case of data loss, permanent loss can occur if there is no backup file to recover the data [22,23]. Moreover, as the EMR-sharing process is performed by a third-party data center or cloud service provider, personal information of the patient can be exposed [24-26]. An EMR contains personal information that can identify the patient. Therefore, privacy issues can arise if sensitive information regarding, for example, abortion clinic visits or records of treatment for sexually transmitted disease, is leaked.

Medical information that directly affects a patient's health must have integrity and be reliable. Moreover, the patient's privacy should be protected from exposure to unauthorized users. Therefore, it is necessary to develop a secure EMR-sharing system that can provide the integrity and reliability of an EMR and protect patient privacy by addressing the problems of the existing centralized EMR-sharing systems. Decentralization of the system has been proposed to complement the problems of the existing EMR-sharing system, and blockchain is receiving much attention as a technology suited for this purpose [27,28].

Blockchain stores data using a shared ledger maintained and managed through consensus by nodes participating in a blockchain network. By storing the previous block's hash value created using an irreversible hash function in the newly created block, blocks form a chain structure in which they are sequentially connected [29,30]. Furthermore, because the data stored on the blockchain cannot be arbitrarily modified or deleted, the blockchain provides strong tamper-resistant performance. Because of these technical characteristics, blockchain provides transparency and integrity of data and enables transactions among users without central administrators and third parties [31]. In addition, the blockchain technology that provides data integrity and transparency through a distributed shared ledger can be made scalable by applying automation technologies such as smart contracts. A smart contract is a digital contract written in code and executed automatically, first devised by Szabo [32]. Since then, smart contracts have been used for digital asset trading on Ethereum, a blockchain platform developed by Buterin [33,34]. By using smart contract technology, users can authenticate the contents of a transaction without the intervention of a third party and can be guaranteed an accurate and automated contract by means of a prewritten code. As blockchain has been applied to various fields, the role of smart contracts has also diversified. When smart contracts are applied to the medical field, various medical services such as remote patient monitoring, clinical trials, and drug supply chain management can be automated [35-37]. Moreover, it is possible to control access rights by using smart contracts so that only users who meet access policies (APs) can access medical data. However, blockchain technology is still at the prototype level, lacks technical stability, and suffers from limited performance, including low throughput and high latency [28,31]. In addition, there are some problems that arise when applying blockchain technology to the EMR system. For example, in the process of propagating transactions to nodes to store data on the blockchain, patient information may be disclosed to multiple users if the EMR is not encrypted. Furthermore, because of the limited block capacity, large-capacity data (eg, medical images) cannot be shared, and there are insufficient measures in place to ensure the patient's ownership of the EMR. Therefore, to apply blockchain technology to the EMR-sharing system, measures to resolve the aforementioned problems are required. Because of these issues, analysis studies are being conducted on whether it is appropriate to apply blockchain technology to the EMR-sharing system [38]. Figure 1 is a flowchart adapted from the study by Wüst and Gervais [38] to determine whether blockchain is an appropriate solution to the problems of the existing centralized database. To share an EMR, the system needs a shared database. However, if a TTP such as a certificate authority (CA) supervising the entire sharing process is semitrusted, privacy concerns may arise. Therefore, there is a limitation to the use of a TTP for EMR-sharing systems. In our proposed system, there are many patients as well as physicians who write EMRs. They should be identified for the purpose of sharing certain

patient information with the appropriate physicians, but they may have concerns about one another’s privacy. Hence, for the EMR-sharing system, permissioned blockchain can be applied.

The blockchain-based decentralized EMR-sharing system has the opposite characteristics to the existing client–server-based centralized system. Through these opposite characteristics, the blockchain-based EMR-sharing system overcomes the current system’s problems and provides various advantages. Unlike the existing centralized system, the blockchain-based decentralized EMR-sharing system exhibits strong resistance to the single point of failure because no central administrator or server controls the system. As multiple nodes operate the decentralized EMR system, data loss or service failure can be prevented even if a specific node fails. Therefore, it is possible to build a more robust system and provide stable service.

In terms of performance, the blockchain-based EMR-sharing system exhibits low throughput and high latency compared with centralized systems because of the data propagation delay between nodes and the consensus mechanism. However, performance problems can be overcome through various

methods, including automation of the system by using smart contracts, lightweight consensus mechanisms, and private blockchain models. In a centralized system, only the central administrator manages the database. Hence, the data stored in the database can be arbitrarily modified or deleted only by the central administrator. However, to modify stored data in the decentralized EMR-sharing system, the consent of most of the blockchain nodes is required; therefore, an arbitrary user cannot modify the data at will. Therefore, the blockchain-based EMR-sharing system provides high data integrity and a transparent process, allowing EMRs to be shared without the intervention of a third party, unlike centralized systems. The decentralized EMR-sharing system prevents data leakage and privacy threats from third parties with these characteristics.

Despite many concerns about technological limitations and suitability, many researchers are studying blockchain-based EMR-sharing systems to take advantage of the benefits of blockchain technology [39-42]. Table 1 shows the differences between a blockchain-based distributed EMR-sharing system and a client–server-based centralized EMR-sharing system.

Figure 1. Decision-making flowchart to determine whether blockchain is an appropriate technical solution to a problem, adapted from the study by Wüst and Gervais [38]. CA: certificate authority.

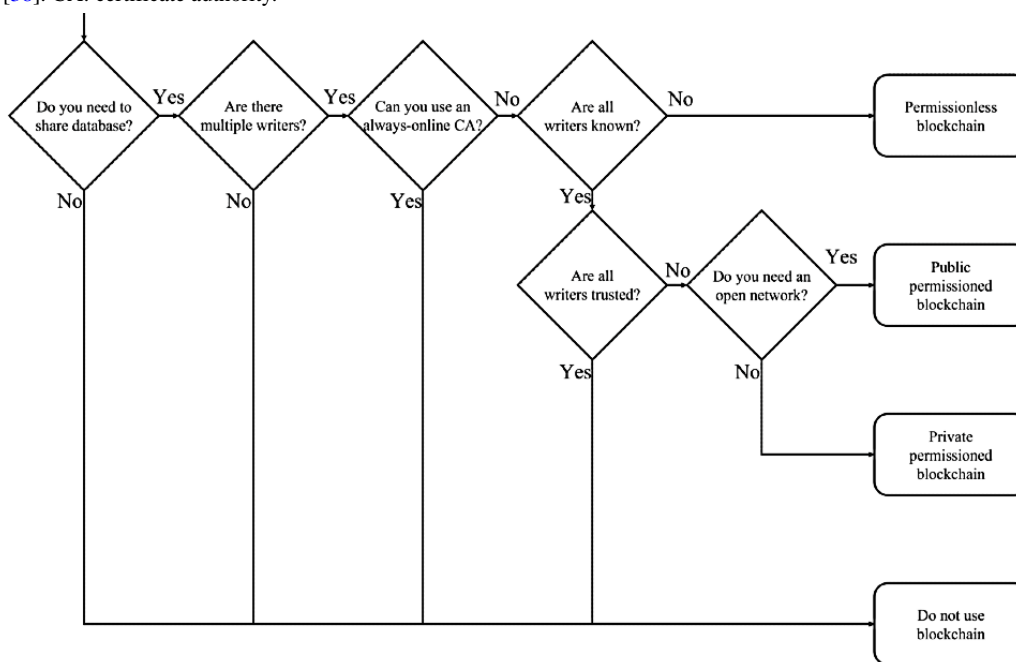


Table 1. Comparison of the decentralized (blockchain) and centralized (client–server) electronic medical record–sharing system.

Characteristics	Decentralized system	Centralized system
System-fault tolerance	Strong	Weak
Throughput	Low	High
Latency	High	Low
Data integrity	High	Medium
Trusted third party	No	Yes
Storage	Distributed ledger	Centralized database
Privacy preservation	Strong	Weak

Related Works: Blockchain Technology in Medical Care Fields

Researchers have proposed various EMR-sharing system models based on blockchain to secure the integrity and reliability of EMRs and build a secure EMR-sharing environment. The studies on EMR-sharing systems based on blockchain technology are presented in [Table 2](#).

Azaria et al [43] proposed MedRec, a decentralized medical record management system based on Ethereum smart contracts. MedRec manages access rights to medical records using smart

contracts and permissions stored in the blockchain. When a client sends a query request, the gatekeeper checks the client's signature and the blockchain contracts to verify access rights. However, because the system proposed by the authors does not encrypt medical data, there is a risk of leakage of personal information and data during the data-sharing process. Besides, when data are shared, the transaction-processing efficiency is degraded because of the additional processing time required as it is necessary for the request queries to be sent from the provider's local database.

Table 2. Blockchain-based electronic medical record (EMR)-sharing systems.

Year	Authors	Description	Limitation	Entities
2016	Azaria et al [43]	<ol style="list-style-type: none"> The authors proposed a new distributed record management system that handles EMRs Researchers and public health authorities participate in the blockchain network as miners Miners given access to anonymized aggregate data as mining rewards through proof of work 	Scalability and security	Patient and provider
2018	Griggs et al [35]	<ol style="list-style-type: none"> All events between patients and physicians are stored and managed using a customized smart contract in the blockchain All sensor data captured by IoT^a devices are stored and managed in the blockchain Smart devices can provide automated alerts using smart contracts to users and health care providers 	Scalability and security	Patient and hospital
2018	Uddin et al [44]	<ol style="list-style-type: none"> Design a lightweight blockchain model and an encryption algorithm for the IoT-based remote patient-monitoring system 	Centralization, verification cost, and scalability	Patient, IoT device, cloud service provider, and hospital
2018	Maslove et al [45]	<ol style="list-style-type: none"> The authors presented a proof-of-concept blockchain-based clinical trial data management solution, enabling patients and researchers to participate in clinical research 	Scalability and security	Patient and researcher
2019	Guo et al [46]	<ol style="list-style-type: none"> The study presents an attribute-based encryption system for authorization and dynamic authentication of medical on-demand services in remote medical systems Data index management using blockchain for data security of public cloud-based telemedicine services 	Centralization and security	Patients, hospital, cloud service provider, and authorities
2019	Hylock and Zeng [47]	<ol style="list-style-type: none"> The authors proposed a proxy re-encryption-based redactable blockchain system for a privacy-preserving and efficient medical data exchange system 	Scalability	Patient, hospital, and researcher
2019	Wu and Du [48]	<ol style="list-style-type: none"> Data-masking techniques were presented to prevent personal information leakage in blockchain-based medical systems IPFS^b, a distributed file-sharing protocol, was used to share large-capacity data such as medical images 	Security	Patient and physician
2020	Abdellatif et al [49]	<ol style="list-style-type: none"> The authors proposed a system model and priority-based data-sharing algorithm using blockchain and edge computing for remote health care systems 	Scalability, security, and centralization	Patient and hospital

^aIoT: Internet of Things.

^bIPFS: InterPlanetary File System.

Hylock and Zeng [47] proposed HealthChain to enhance patient engagement and security in blockchain-based health information exchange systems. Proxy re-encryption (PRE) [50-54] technology was used to prevent leakage of patient private keys and medical data. Furthermore, the authors introduced redactable patient blocks with chameleon hashing to solve the data

fragmentation problem and reduce storage and computation overhead by modifying the data. However, in the system proposed by the authors, there is a fatal problem: patients must share their private keys with an external third party for re-encryption. Moreover, for patients to share and manage their medical data, they must continuously participate in the

blockchain network, burdening patients who have limited resources, unlike hospitals and research institutes.

Wu and Du [48] proposed an EMR security-sharing model based on blockchain for improving privacy and data scalability in medical data-sharing systems. An EMR security-sharing model based on blockchain uses data-masking technology to hide sensitive information stored in the medical data to prevent leakage of personal information. Moreover, the InterPlanetary File System (IPFS) [55], a distributed file-sharing protocol, was used to overcome the difficulty of sharing medical data because of the limited block size. However, depending on the masking level, the privacy protection offered by data-masking technology varies in performance. The problem is that applying excessive masking makes it challenging to use the required information, and when the level of masking is low, specific values can be tracked and predicted.

Abdellatif et al [49] proposed ssHealth, a smart and secure health care system, which is a distributed health care system that enables convenient medical data-sharing among various institutions using blockchain and edge computing. The ssHealth system divides medical data processing, access control, and data sharing into local and blockchain networks and presents a data-sharing security algorithm based on the importance of data to enable safe medical data-sharing. However, there is a problem: it is not possible to guarantee stable service quality because of differences in validation time, depending on the security level. Besides, there is a risk of centralization and privacy breaches because unencrypted medical data and patient personal information pass through edge nodes in the local network.

Existing studies on blockchain-based EMR-sharing systems have used blockchain models designed for cryptocurrencies such as Bitcoin and Ethereum. However, existing blockchain models for cryptocurrency have limitations in providing the security and scalability required in sharing EMRs. The system also failed to meet the requirements of EMR-sharing systems as defined in *Connecting Health and Care for the Nation: A Shared Nationwide Interoperability Roadmap – Version 1.0*, prepared by the Office of the National Coordinator for Health Information Technology (ONC) [56]. Therefore, it is necessary to develop a blockchain-based EMR-sharing system that overcomes the limitations of existing systems and satisfies the security framework defined by the ONC.

In this paper, we propose a patient information exchange (PIE) system. The proposed blockchain-based EMR-sharing system overcomes the limitations of existing blockchain-based EMR-sharing systems and satisfies the privacy and security framework defined by the ONC. Furthermore, our proposed system prevents data loss and privacy breaches in sharing data through the data encryption scheme based on re-encryption, ensuring strong data security. Moreover, data integrity is ensured by preventing the forgery and alteration of EMRs by using the decentralized ledger structure and the unique hash value of the data. Furthermore, allowing patients to set their data-access rights ensures patient ownership of their EMR and establishes a patient-centered medical system. Moreover, the PIE system provides improved performance by solving the low processing

performance and scalability issues due to the limited block capacity of the existing blockchain through the distributed data-sharing method using the IPFS. As a result, we contribute business process optimization, cost reduction, patient outcome improvements, and enhanced compliance in the health care field [57,58].

Methods

System Model

Here, we describe the proposed PIE system. In the *Components of the Proposed PIE System* section, we define the entities that make up the components of the system and describe each entity's role. In the *EMR Transaction Structure* section, we describe the structure and components designed to share EMRs effectively. Finally, in the *Security Levels of EMRs* section, we discuss the security level, which depends on the EMR data type and classifies the data based on the type. The system model of the proposed PIE system is shown in [Figure 2](#).

We propose a blockchain-based PIE system to improve the security and efficiency of the EMR-sharing process. To prevent forgery of EMRs and protect patient privacy, we use a consortium blockchain model in which only authorized users can participate. The medical consortium that operates and manages the blockchain comprises state-approved and trusted medical institutions. As the proposed blockchain-based PIE system uses a private blockchain model, the consensus algorithm in the block generation process is not addressed. Instead, the chain is constructed by sequentially storing the generated EMR transactions to create a block and connecting them. Hospitals and medical institutions serve as blockchain nodes that issue EMR transactions and store them in block form. Health care workers and patients who create and use EMRs participate in the blockchain network as users by using IDs issued according to user type after a certification process by a CA. Users participating in the blockchain network can register their EMRs on the blockchain and use them at any time. The proposed PIE system is a patient-centered EMR-sharing system where patients directly participate in the EMR upload and EMR-sharing process. The patient directly generates a key to encrypt the EMR and defines the categories of users who can access the EMR. By allowing patients to manage their own EMRs, we build a user-centered system that protects patients' privacy and gives them ownership of their EMRs.

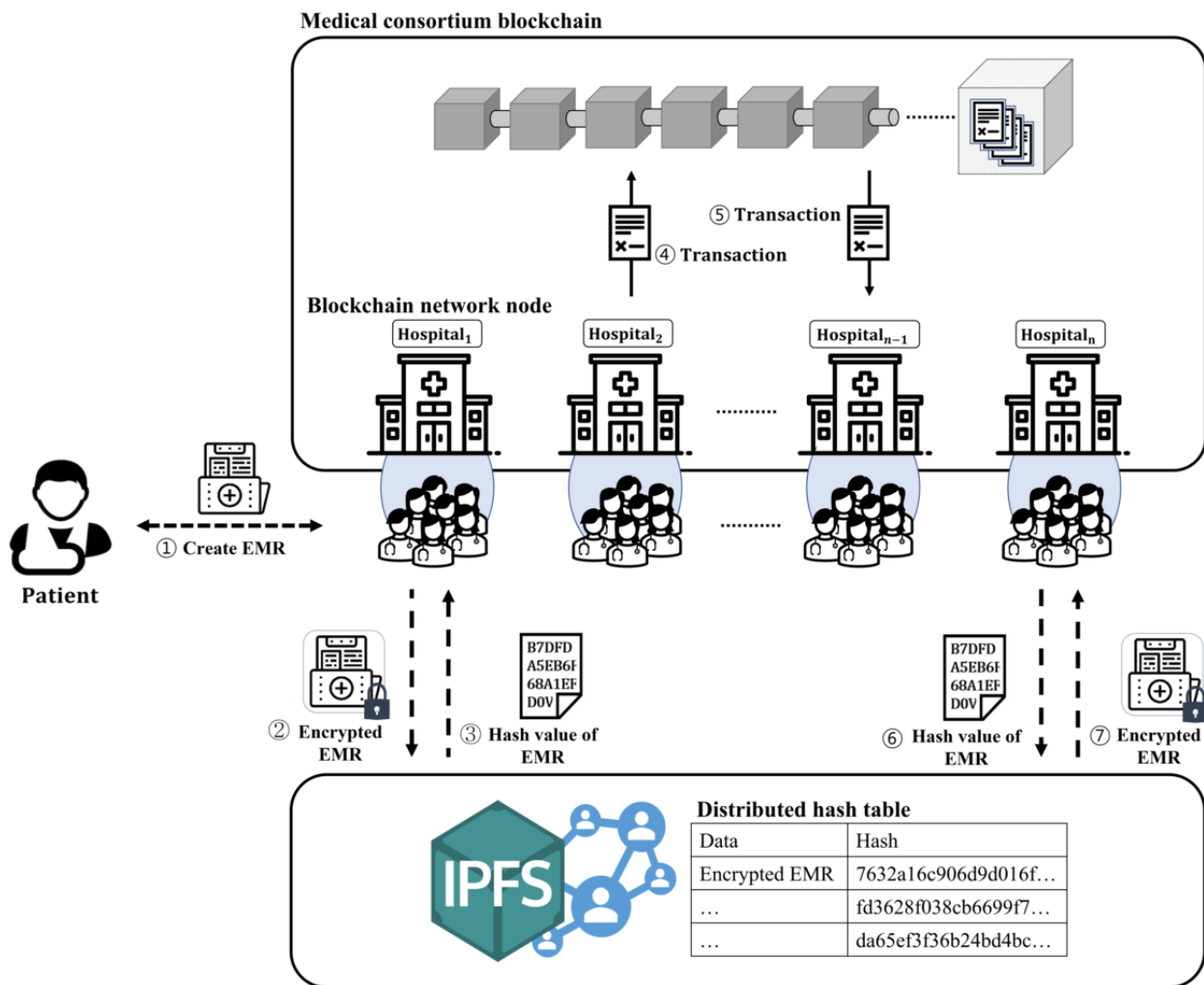
The proposed PIE system securely protects patient EMRs from security threats such as data forgery and personal information leakage, which can occur during EMR management and sharing. To protect EMRs from the aforementioned security threats, we use public key-based asymmetric encryption and our proposed PRE-based decryption authority delegation mechanism. The proposed decryption authority delegation mechanism prevents private key leakage by decrypting data encrypted with the public key. Moreover, delegating authority to decrypt data solves the problem of data access in an emergency when a patient cannot respond to a request for access to their EMR, as de Oliveira et al [59] suggested. The proposed PIE system provides a re-encryption key that enables the physician who created the EMR to act on behalf of the patient in an emergency when the

patient is unable to control access rights to the EMR. The re-encryption key re-encrypts the EMR encrypted with the patient's encryption key into a form that the physician can decrypt with the physician's private key.

The performance and scalability of the PIE system are enhanced by using the IPFS, which supports distributed data-sharing technology. The EMR encrypted with the patient's encryption key is stored on the IPFS network, and the hash value of the

EMR is stored on the medical blockchain in the form of metadata. Instead of storing the data as a whole in the blockchain, it is possible to reduce the load on the system by storing only the hash value of the data. Furthermore, if data are shared using the IPFS, large-capacity data such as magnetic resonance imaging, computed tomography, and endoscopy images can also be shared, improving the scalability of the blockchain system.

Figure 2. The proposed blockchain-based patient information exchange system model. EMR: electronic medical record. IPFS: InterPlanetary File System.



Components of the Proposed PIE System

The proposed PIE system consists of blockchain nodes (medical consortium), users of the blockchain network (patients and health care workers), and the IPFS. The role of each entity is outlined in the following paragraph:

A medical consortium consisting of hospitals and medical institutions that wish to share EMRs builds and manages a distributed ledger as operator of a permissioned blockchain network in which only authorized users can participate. The medical consortium blockchain stores the information of the EMRs generated by each hospital. The information recorded on the blockchain is a hash value of real medical data stored in the IPFS and simplified medical information that users can comprehend. Data registered on the blockchain cannot be

arbitrarily deleted or modified, providing high reliability and medical data integrity. Patients and physicians, who are the users of the blockchain network, share EMR information through the network. Patients can use a decentralized app to share their EMRs in the PIE system. Furthermore, patients set their APs for their EMRs and generate re-encryption keys for re-encryption. Unlike traditional hospital-centered health care systems, the PIE system guarantees the patient's ownership of their EMR. In a patient-centered health care system, where patients have rights to their own EMRs, they have the freedom to choose who can use their EMR and their data at any time. Furthermore, patients may sell their medical data to research institutions or hospitals, in addition to using the data for therapeutic purposes. Health care workers consist of reliable physicians and health care service providers such as medical researchers and insurance agents. Health care workers use

computer systems at hospitals or medical institutions to encrypt EMRs generated during the patient treatment process and upload them to the IPFS. After uploading the EMRs, health care workers submit the EMR information to their hospitals and institutions. Health care workers also serve as consumers of medical data by, for example, sharing EMRs through a blockchain network to treat patients or using the data for clinical research. The IPFS is a distributed file-sharing system that splits data stored on multiple computers worldwide into small pieces and shares only a portion. The distributed data-sharing method used by the IPFS enables rapid sharing of large-capacity data such as magnetic resonance imaging or computed tomography images. In addition, the IPFS prevents duplicate creation and storage of medical data by managing data with hash values based on data content.

Threat Model

In this study, we consider the traditional cryptographic system, not the postquantum cryptographic system. Therefore, we use the discrete logarithm problem, which is one of the difficult problems of 1-way functions. The discrete logarithm problem is one where given $x, y \in Z_q^*$, it is difficult for any probabilistic

polynomial time attacker A to find a value $m \in Z_q^*$ such that $x = y^m$. Therefore, the attacker cannot obtain the private key from a public key or ciphertext. Our system model considers external threats from outside the system and internal threats from the system participants. We assume that both threats are in the form of a logical attack, not a physical attack. The external threats target the patients' private data such as the EMR, insurance details, and other personal information. For example, an external attacker wants to eavesdrop on all communication among the participants to obtain patients' personal information. Internal attackers can include health care researchers or insurance agents. They are allowed access to limited information in the form of an abstract regarding disease and length of hospital stay, not details of the disease or patients' personal information. However, internal attackers are snooping for patients' private data; therefore, they try to access their medical information. In addition, internal attackers attempt to manipulate clinical results or commit insurance fraud by arbitrarily forging a patient's EMR. Table 3 shows the attack scenarios and threat situations considered in the proposed blockchain-based EMR-sharing system.

Table 3. Attack scenarios and threats considered by the proposed system.

Types and attack scenario	Threats
External threats	
Eavesdropping	Private data leakage (eg, electronic medical record and personal information)
Denial of service	Service unavailable
Internal threats	
Abnormal access	Private data leakage
Data forgery	Unexpected output

EMR Transaction Structure

The proposed blockchain-based medical system uses transactions designed to effectively share the desired medical data while preventing leakage of personal information and data when uploading the medical data to the blockchain. A unique identifier or ID is used in the blockchain network by the physician who created the EMR and the patient who is the owner of the generated data. The CA issues a user ID according to the type of user participating in the blockchain network. A user ID is a randomly generated value consisting of numbers and letters; it is possible to identify users but not know who the owner is. Information about users who can map users to user IDs is securely managed by a CA such as the trusted government authority that issued the ID. As the user IDs are correlated, users are protected from the threat of personal information leakage [60,61]. The timestamps record the time the transaction was created. Medical information contains minimal necessary

medical information, excluding sensitive information that can identify the user from the patient's EMR. For example, even if information such as the gender of the patient, type of disease, age, and exercise status is disclosed, it is not a serious problem because the owner of the data cannot be identified. The information is only used in the search process to identify a specific EMR of interest among the various EMRs stored in the blockchain. The metadata contain the hash value, which is the address value of the data received after uploading the encrypted EMR to the IPFS. Using this, a specific EMR can be shared in the IPFS. The contract code contains the code to execute the smart contract, for example, the user's AP. The signature is the one created by using the private key of the physician who created the transaction (the physician who treated the patient and generated the EMR). The structure of transactions for effective and safe EMR management and sharing is shown in Table 4.

Table 4. Transaction structure for electronic medical record (EMR) sharing.

Field	Definition
User ID	IDs of the patient and physician
Timestamp	Time the transaction was created
EMR information	Summary of information in the EMR
Metadata	Hash value of encrypted EMR
Contract code	Patient's defined access permission policy
Signature	Signing with the user's private key

Security Levels of EMRs

Patient EMRs may include data relating to clinical trials and insurance as well as sensor data generated by health care devices, in addition to medical information generated during the process of receiving treatment in hospitals. Depending on the EMR data type, the required security levels will differ. For example, if information such as name, residence, and social security number, which can identify an individual, is leaked to outside parties, it can lead to serious personal information leakage; consequently, a high security level is required. Conversely, information that is not personally identifiable, such as gender, age, eating habits, and exercise status, does not

require a high security level because it is not a serious problem even if this information is disclosed to outside parties. Therefore, it is necessary to provide differentiated security levels and separate management, with respect to the sensitivity of the personal information, according to the EMR data type.

The minimum security level required for each data type is established by categorizing privacy sensitivity according to data type and evaluating accessibility and data potential per user. The security levels assigned according to the sensitivity of private information fall into three classes: private, moderate, and low. [Table 5](#) lists the security levels differentiated according to the type of information contained in the EMR.

Table 5. Security levels required depending on the type of information contained in the electronic medical record.

Division and class	Security level
Medical record	
Medical information	Private
Admission record	Private
Prescription	Private
Medical imaging (x-ray, magnetic resonance imaging, and computed tomography)	Private
Clinical trial	
Medical device	Low
Medicine	Low
Clinical observation	Low
Omics (genomics)	Low
Lifelog	
Sensor data (weight, heart rate, and sleep pattern)	Moderate

EMR-Sharing Process in the PIE System

Overview

This section presents the EMR-sharing process that protects the patient's EMR from various attacks and safely shares it. Moreover, it describes the work performed in each process. A more detailed description of each process-specific algorithm is provided in [Multimedia Appendix 1](#). The proposed EMR-sharing process consists of a user registration phase, an EMR upload phase, and an EMR-sharing phase. The user registration phase

concerns joining the blockchain network so that users such as patients and physicians can manage and share medical data. The EMR upload session concerns registering on the blockchain the medical data generated when treating patients. The process of publishing a patient's EMR on the blockchain involves data encryption, access rights setting, transaction creation, and uploading. Finally, the EMR-sharing phase concerns the process of downloading the encrypted patient EMR from the blockchain, re-encrypting it, and then decrypting it to obtain the original version. The notations used for EMR encryption and re-encryption in the proposed system are listed in [Table 6](#).

Table 6. Notations used for electronic medical record (EMR) encryption and re-encryption in the proposed system.

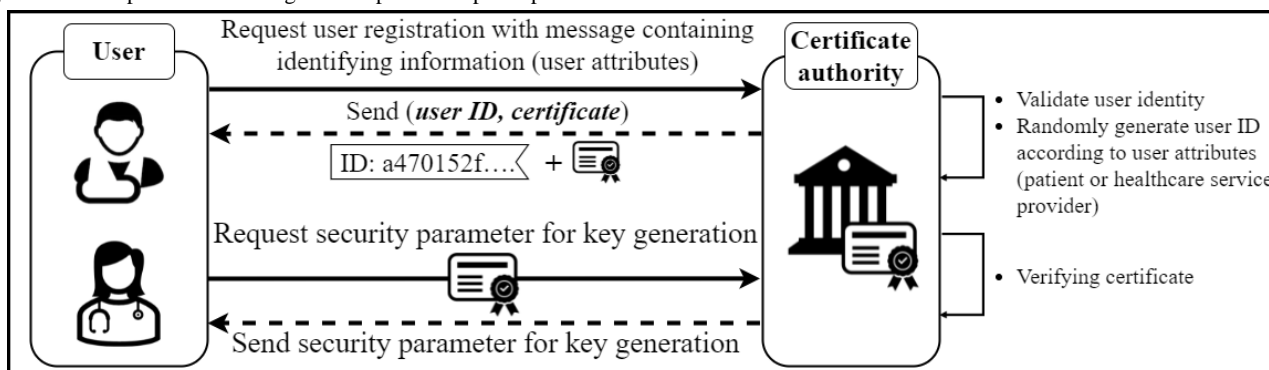
Notation	Description
ID	User ID (patient or physician)
SK	Private key of the user
PK	Public key of the user
DEK	Dedicated encryption key to encrypt EMR
RK	Re-encryption key
C_i	Encrypted EMR
P_i	Patient _i
D_i	Doctor _i
AP	Access policy
hash _i	Hash value of the encrypted EMR

Initial Phase: User Registration

Patients and physicians want to participate in the blockchain network to manage EMRs and securely share them with other users. Users participate in the blockchain network through the user registration process, which consists of an identity registration phase to register the user’s identity and an authentication phase to obtain security parameters to generate an encryption key. Figure 3 shows the user registration process for users to join the blockchain network. In the ID registration step, the user sends the ID registration request message that contains the user’s attributes to a CA. The user attributes identify whether the user is a health worker or a patient. The CA also uses the user attributes to identify the user’s ID. Once the identification is complete, if the user is legitimate, the CA classifies the user type based on the user attributes. Subsequently, depending on the user type, a user ID is generated and delivered to the user along with the certificate. Users who are judged to be not legitimate because their attributes are not validated will be denied user registration. A user who has successfully registered an account then sends a message, including the certificate, to the CA requesting security

parameters to generate the encryption key needed for EMR sharing. If the certificate is valid, the CA provides the security parameters to the user. Thus, through the user registration step, users who join the blockchain network generate a public key and a private key using the security parameters received from the CA. First, the user selects a random decimal number corresponding to $x \in Z_q^*$. The selected x is set as the user’s secret key and is never shared. Next, users generate a public key for use in the network using their private key and the key generator. Users who can directly generate the encryption key that they use for EMR sharing can generate an encryption key each time an EMR is created, thus protecting the EMR with a different key each time. Because of the nature of an EMR that has been used for a long time, the encryption key must be maintained as is when data are encrypted [62]. However, this introduces security vulnerabilities, endangering EMR and patient privacy. To prevent this, in our system, users create the encryption key themselves, and the existing EMR uploaded to the IPFS is updated with data encrypted with the new key through a data version update. Users who successfully register an account and generate an encryption key can share medical data through the blockchain network.

Figure 3. Initial phase: the user registration process to participate in the blockchain network.



Phase 1: EMR Upload—Block Generation

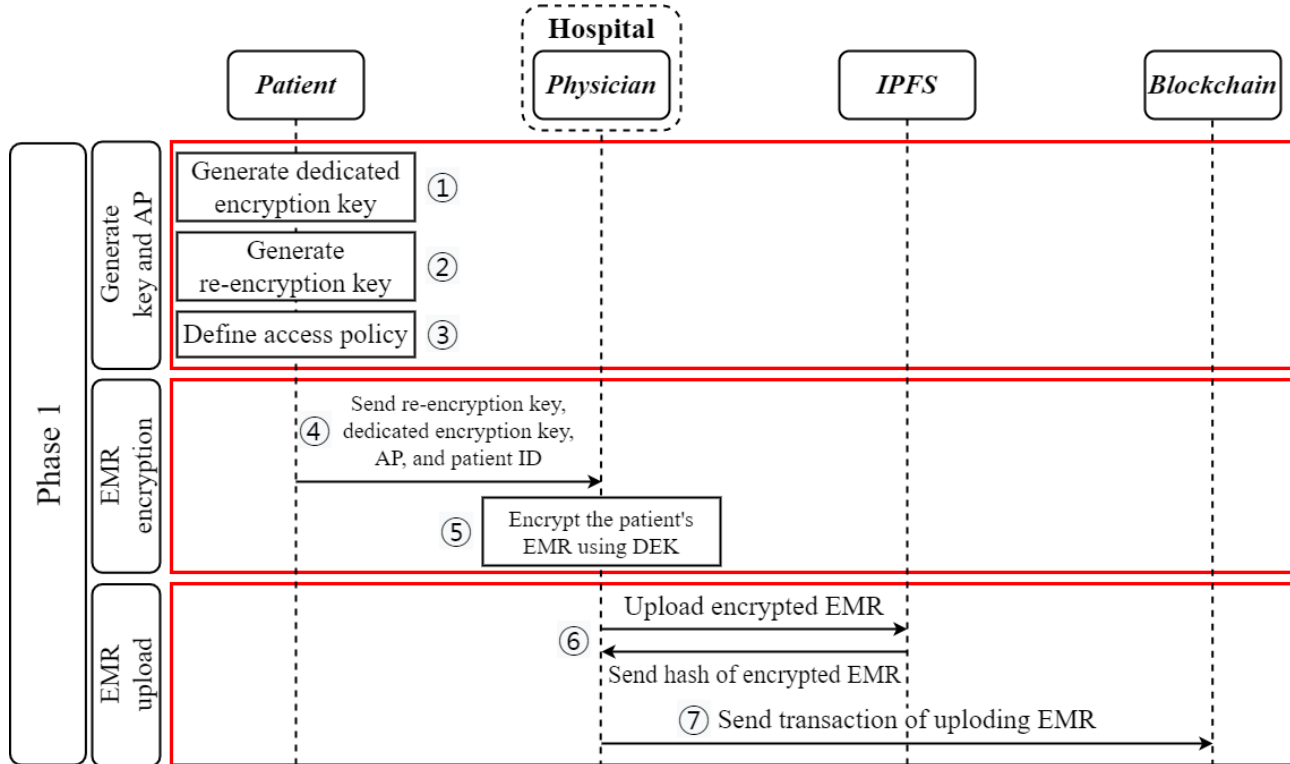
When a medical record is generated for treatment, the patient creates a dedicated encryption key to encrypt their EMR and a re-encryption key RK_{PD_i} to re-encrypt the encrypted EMR. The re-encryption key is generated using the patient’s private key

and the physician’s public key. At the same time, the patient creates an AP that defines the level of users who can access the EMR. Next, the patient sends the dedicated encryption key RK_{PD_i} , the AP, and the user ID, which was created by the patient, to the physician who provided treatment. Upon receiving the

message from the patient, the physician encrypts the patient’s EMR with the patient’s dedicated encryption key. The physician then uploads the encrypted EMR, C_p , to the IPFS and receives the hash value, $hash(C_p)$, of the data the IPFS has stored. Next, the physician submits the user ID related to the EMR, the minimum information required to distinguish EMR_p , and

$hash(C_p)$ to the medical institution to which they belong. Finally, the medical institution uses the information in the message to create a transaction and publish it on the blockchain. Figure 4 shows the EMR upload flow diagram of the proposed blockchain-based PIE system. The method of uploading EMR information to the blockchain follows algorithm 1 defined in Textbox 1.

Figure 4. Phase 1: electronic medical record (EMR) upload flowchart of the proposed blockchain-based patient information exchange system. AP: access policy; DEK: dedicated encryption key; IPFS: InterPlanetary File System.



Textbox 1. Algorithm 1: the electronic medical record upload.

Algorithm 1

- Input: Secret key_{Patient}, public key_{Doctor}, dedicated encryption key, user ID_{Patient}, hash(C_{Patient}), and access policy
- Output: Re-encryption key, access policy, summary information from electronic medical record, hash value returned by the InterPlanetary File System, and transaction
 - The patient selects a random security parameter value r to generate a dedicated encryption key for encrypting their electronic medical record.
 - The patient uses their secret key SK_P and the physician’s public key PK_D to generate a re-encryption key $RK_{P \rightarrow D_i}$ for re-encrypting their EMR_p .
 - The patient generates an access policy that defines which users can access their electronic medical record.
 - The patient transmits the dedicated encryption key, $RK_{P \rightarrow D_i}$, access policy, and $UserID_P$ to the physician who treated them.
 - The physician who receives the dedicated encryption key, $RK_{P \rightarrow D_i}$, access policy, and $UserID_P$ from the patient encrypts the patient’s medical record EMR_p using the dedicated encryption key.
 - The physician uploads the encrypted patient electronic medical record C_p to the InterPlanetary File System and receives the $hash(C_p)$, the hash value of the electronic medical record.
 - The physician submits the patient’s ID, summary information from EMR_p , and hash value to the hospital.
 - The hospital uses the received information to create a transaction and uploads it to the blockchain network.

Phase 2: EMR Sharing

Users who want to share and use a particular EMR can search for it on the blockchain through a smart contract and request a re-encryption key. As the medical field is closely related to human life, the target and purpose regarding the EMR data must be legitimate. However, it would be difficult to find the desired EMR among countless data because it is hidden to protect the patient's privacy. Therefore, in the proposed PIE system, smart contracts are applied so that patients and health care workers who are users of the blockchain network can perform a quick and accurate search for the data they want. Moreover, the efficiency of the EMR-sharing process has been improved by automating the process of requesting decryption rights after searching for an EMR. The requester uses summary information from the EMR and the user ID (the ID of the hospital that uploaded the data or the patient's ID) to quickly search for a transaction containing the information of the desired EMR. Then, the requester downloads the encrypted EMR from the IPFS using the acquired transaction information. As the downloaded EMR is encrypted with a dedicated encryption key, it must be decrypted using the patient's private key or re-encrypted using the re-encryption key before using it. However, because sharing the patient's private key is very dangerous, the requester must send a message to the patient requesting a re-encryption key for re-encryption. The message

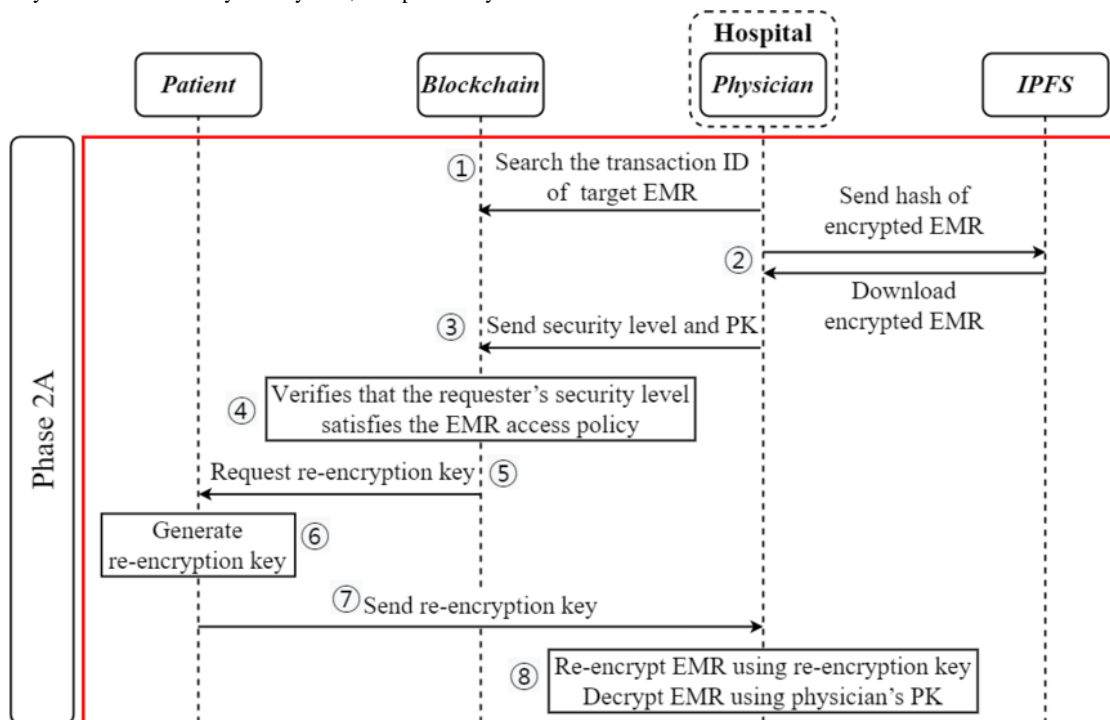
requesting the re-encryption key goes through the user verification process of the access check contract. Initially, the user verification process checks whether the requester's security level satisfies the patient's EMR AP. When it is confirmed that the user has met the required security level, a message requesting a re-encryption key is sent to the patient, the owner of the EMR. Upon receiving the message requesting the re-encryption key, the patient sends $RK_{P \rightarrow Requester}$ to the requester. If a patient cannot issue a re-encryption key because they have been incapacitated by a serious illness such as acute stroke, the physician participating in the EMR generation can temporarily issue a re-encryption key according to the emergency event procedure. The requester who receives $RK_{P \rightarrow Requester}$ can re-encrypt C_P and then decrypt with their private key. The EMR-sharing procedure is performed in the order specified in [Textbox 2](#). The EMR-sharing flowchart of the proposed blockchain-based PIE system is illustrated in [Figure 5](#). The biggest advantage of using PRE technology for medical data security is that users can decrypt a downloaded EMR encrypted with their key without the patient's private key. Therefore, it minimizes the threat of leakage of the patient's private key and information. Moreover, the proposed re-encryption technology-based medical data encryption method can satisfy the medical field's requirements for sharing medical data while protecting the medical data from other eyes when sharing the EMR. [Figure 5](#) shows the EMR-sharing procedure in the proposed PIE system.

Textbox 2. Algorithm 2: electronic medical record sharing.

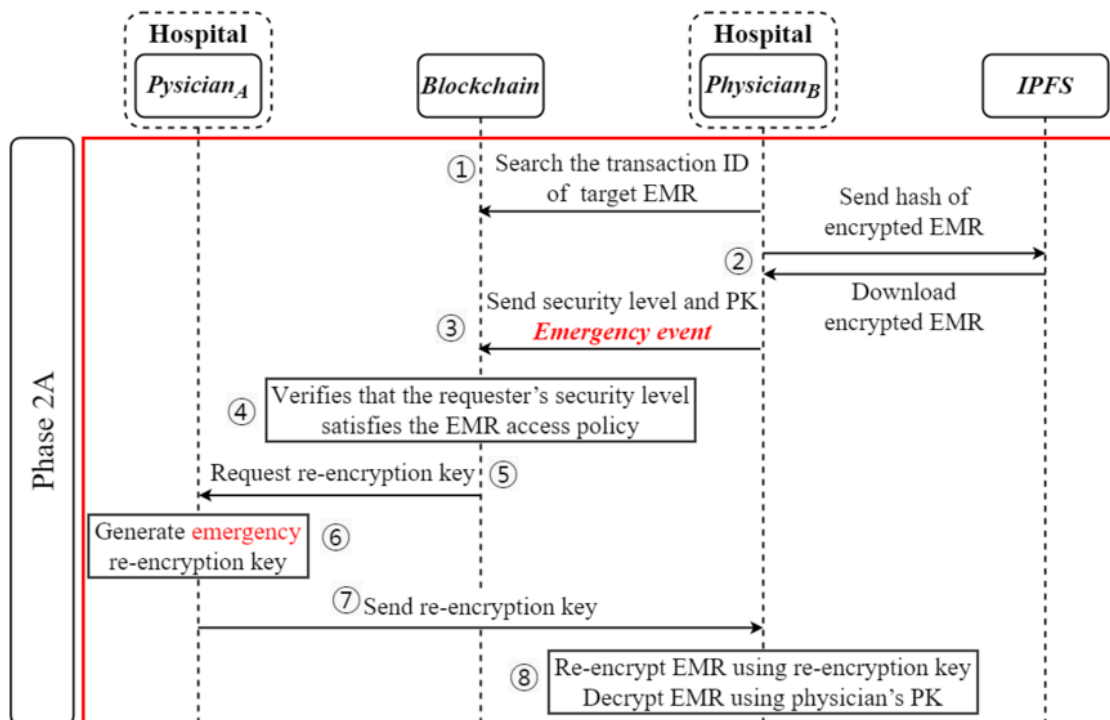
Algorithm 2

- Input: Summary information from the electronic medical record, user ID, and user's security level
- Output: Re-encrypted electronic medical record C_d and $RK_{P \rightarrow Requester}$
 - The physician executes a smart contract for electronic medical record retrieval to find a transaction containing the information of the desired electronic medical record.
 - The smart contract uses the user ID or the electronic medical record's summarized information to find a transaction containing the desired information and returns it.
 - The physician who receives the transaction information downloads the encrypted electronic medical record C_P from the InterPlanetary File System using the hash value contained in the transaction.
 - The physician executes the re-encryption key request smart contract to request the key RK_{EMRP} for re-encryption from the patient who is the owner of the encrypted electronic medical record.
 - The smart contract for requesting the re-encryption key performs a user authentication step to verify that the security level submitted by the user who requested the message transmission satisfies the access policy set by the patient.
 - If the security level of the user requesting the re-encryption key satisfies the access policy, the smart contract sends a message requesting the re-encryption key to the patient (if the user's security level does not meet the criteria, the request is denied).
 - The patient who receives the message requesting the re-encryption key generates $RK_{P \rightarrow Requester}$ using the requester's public key and send it to the requester.
 - The physician receiving $RK_{P \rightarrow Requester}$ uses it to re-encrypt C_P into C_D .
 - The physician uses their private key to decrypt C_D to obtain the original EMR_P .

Figure 5. Phase 2: electronic medical record (EMR)–sharing process flowchart of the patient information exchange system. (A) EMR-sharing process in the general case where the patient controls the re-encryption key, (B) EMR-sharing process in emergencies where the patient has no control over the re-encryption key IPFS: InterPlanetary File System; PK: public key.



(A) EMR-sharing process in the general case where the patient controls the re-encryption key



(B) EMR-sharing process in emergencies where the patient has no control over the re-encryption key

Results

Simulation Design

A simulation was designed to verify that the proposed blockchain-based medical system sufficiently reflects the medical field’s requirements and enables safe data sharing. We simulated the process of sharing encrypted EMRs over a

blockchain network. In the EMR-sharing process of uploading and downloading EMRs over the network, we checked the effect of the time taken to process the data and the size of the data shared on the system’s performance. Moreover, the performance of the proposed smart contract–based re-encryption key–sharing method was verified by measuring the execution time of the smart contract. This paper does not cover improvements to the

consensus process performed on the blockchain network; therefore, the improvements were not evaluated.

The test environment was designed based on the data-sharing process defined in the *Methods* section. The simulation consisted of 2 physicians, a patient, a medical consortium, and a public IPFS network participating in the blockchain network as the minimum unit for sharing an EMR. Entities on the network were classified into two categories: a host operating a blockchain network and a guest operating on the network. All processes in the proposed system can be viewed as interactions among guests. In the test environment, there was a consortium to which many hospitals belonged, and each hospital had a Flask server with a different port. The code of the software development kit used to implement the blockchain network is Node.js. The host computer was a PC running Windows 10 Pro 64x (Microsoft Corp) in a wired environment, with 32 GB memory and Intel

Core i7-10700K central processing unit at 3.80 GHz. The guest computer was a PC running Ubuntu 18.0.4 64x (Canonical) in a wired environment, with 10 GB memory and Intel Core i7-10700K central processing unit at 3.80 GHz. The communication rates considered were the download and upload rates in megabits per second: host upload and download rates and guest upload and download rates.

The simulation was performed on 1 PC to directly compare the host's and guest's processing times. Blockchain implementation was performed using Hyperledger 2.3.1 [63]; Apache CouchDB [64] was the state database. The network consisted of 4 orderer nodes, 3 organizations, 2 peer nodes for each organization, and 1 channel. The chain code for smart contracts used Go. [Table 7](#) shows the parameters used for the proposed PIE system simulation.

Table 7. Simulation parameters.

Parameters	Values
Data size	0.4 kB, 1 MB, 10 MB, 100 MB, and 1 GB
Data type	CSV (text) and DICOM ^a (images and videos)
Number of orderer nodes	4
Number of organizations	3
Number of peer nodes	6
Number of channels	1
Data rate	100 Mbps
Block size	1 MB
Block timeout	2 seconds
Database	Apache CouchDB

^aDICOM: Digital Imaging and Communications in Medicine.

Simulation Results

To evaluate the performance of the proposed PIE system, we measured the time required for the EMR-sharing process and the execution time of the smart contract for re-encryption key sharing. The EMR-sharing process was divided into upload and download processes, and the time taken to perform each process was measured. The execution time of the EMR upload process was defined as the time taken to upload the EMR to the IPFS and post the returned EMR hash value to the blockchain. The execution time of the EMR download process was a measure of how long it took users to download the EMR over the IPFS. Considering the characteristics of an EMR that supports various types of data, the simulation was performed using various data, ranging from text format (0.4 kB) to medical images (1 GB). The simulation measured only the time required in the communication process for exchanging data among users and did not consider the impact on the process of the data encryption and decryption operations. To objectively evaluate the performance of the proposed system, we performed a comparative analysis with existing blockchain-based medical information exchange systems. Simulations were performed for three types of systems (an on-chain-based system designed for cryptocurrency, an Ethereum-based system using the IPFS, and

a PIE system); the results of the simulations for the EMR upload process are shown in [Multimedia Appendix 2](#) [47,48].

Through the EMR upload simulation it was confirmed that the larger the data to be uploaded, the longer it takes; the larger the amount of data to be uploaded, the higher the required data rate. As a result, the processing time increased dramatically for data that exceeded the acceptable data rate (100 Mbps) in the simulation environment. In an on-chain-based system that stores data in the original form in blocks, the size of data that can be uploaded is limited to 1 MB, which is the maximum size of the block; therefore, there is no simulation result for data beyond that size. Most of the time taken to upload an EMR was when a query request needed to be made to the blockchain network, which took an average of 2.1 (SD 0.0343) seconds. The actual time taken to upload an EMR to the IPFS increased depending on the size of the data, but it was very short. A graph of the time it takes to upload EMR to IPFS can be found in [Multimedia Appendix 3](#). The time required to upload data that ranged from 0.4 kB to 100 MB was relatively short compared with the query request time; therefore, it did not significantly affect the overall EMR upload time. Again, when uploading data that ranged from 0.4 kB to 100 MB, the overall EMR upload time was comparable with the query request time (average of 2.1 seconds, SD value

= 0.2947). However, when uploading >500 MB of data, the time taken to upload the EMR to the IPFS was longer than the query request time, which affected the overall EMR upload time. Uploading 500 MB and 1 GB of data took an average of 4.5 (SD 0.1329) seconds and 5.7 (SD 0.21) seconds, respectively, for the Ethereum-based system and the proposed PIE system. The PIE system and the Ethereum-based system showed similar performance in that the EMR was distributed and shared using the IPFS. However, in the Ethereum-based system, there is a problem: to decrypt the shared medical data, the patient's private key needs to be shared or the patient needs to directly decrypt the shared medical data. This gives rise to a fatal security problem: the patient's private key and EMR can be leaked directly to others. In contrast, the PIE system prevents the leakage of the patient's private key and EMR by using a re-encryption scheme and enables users who have shared their EMRs to decrypt them smoothly, providing high security along with the same performance as that of the Ethereum-based system. Uploading the actual EMR to the IPFS and sharing it through a decentralized technique has 3 important implications in a blockchain-based medical data system. First, medical data can be shared without capacity limitation through a peer-to-peer network. This advantage can thus alleviate the problem of low processing efficiency and data scalability because of the blockchain's limited block capacity. Second, by storing the hash value, which is the unique address value of the data, it is possible to reduce the blockchain's storage burden and solve the EMR reduced redundancy storage problem. When sharing an EMR, the PIE system reduces the burden on the nodes and allows data to be shared faster. The distributed data-sharing method using the IPFS determines the performance according to the number of users sharing data, showing higher performance as the number of users increases. Therefore, EMR sharing using the distributed data-sharing method is effective for the medical system because it can share data faster while reducing the burden on the node. However, the data-sharing method using the IPFS has a problem in that there must be at least one node that stores the data to be shared in the IPFS network. If the data are not stored (pinned) on the IPFS network, the shared system can fail. To prevent this, the hospital that created the EMR needs countermeasures such as storing data in a local database in preparation for the worst-case scenario after uploading.

The simulation results for downloading the EMR posted on the blockchain are presented in [Multimedia Appendix 4](#) [47,48]. The simulations we conducted used various data sizes ranging from 0.4 kB to 1 GB. The IPFS used in the simulations is an open network, with a variable number of users participating in data-sharing operations. Therefore, even with a blockchain system using the same IPFS, differences in performance may occur depending on the number of users participating in the network. The Ethereum-based blockchain system [48] and the PIE system we propose use the IPFS to alleviate the blockchain's scalability problem. However, there is a big difference between the 2 models in the pieces of information they store. In the model proposed in the study by Wu and Du [48], only the detailed information of the EMR is stored, whereas in the PIE system we propose, the original EMR is stored as is. Storing the original data of the EMR generated in the medical process can ensure the integrity of medical data and

increase its usability in medical systems. The simulation results show the difference in performance between storing only the detailed information of the EMR and storing the original EMR as is. Even considering that the performance of the IPFS may fluctuate depending on the number of participating users, it takes less time to download the original EMR than it takes to download the details of the EMR. These results mean that the proposed PIE system provides higher scalability in that it can provide the original EMR to users more quickly. The average time taken to download 1 MB of medical data encrypted using the unique address value of the medical data uploaded to the IPFS in the proposed PIE system is 0.01014 (SD 0.0028) seconds. This is approximately 5.5 times faster than the average download time of 0.0562 (SD 0.0052) seconds taken by the on-chain-based blockchain model without the IPFS. Existing blockchain systems that publish and share data without using the IPFS increase the burden and have limitations in scalability as the size of the data to be shared increases. However, there is no limit to the size of the EMR posted on the blockchain in the proposed method, ensuring high scalability and higher processing performance than that of existing systems. Therefore, for medical systems that need to share medical data, it is more effective to use the distributed data-sharing method.

We performed a simulation of the smart contract-based re-encryption key-sharing process. The re-encryption key-sharing process verifies the user requesting the re-encryption key to use the patient's EMR and passes the re-encryption key to the user. The user who receives the re-encryption key performs the re-encryption process and finally decrypts the encrypted EMR using their private key to use the patient's EMR. A graph of the smart contract simulation results for re-encryption key sharing is in [Multimedia Appendix 5](#). The simulation was performed with 4000 epochs, and the quarterly smart contract execution time and average time required are presented. As the re-encryption process and decryption process are performed by the user alone, they are not included in the smart contract's execution time for sharing the re-encryption key. The average time taken to verify a user requesting a re-encryption key to use the patient's EMR and grant decryption rights to an authorized user is 3.3543 (SD 0.4959) seconds. The data security process using the re-encryption method effectively protects medical data; in addition, it can provide convenience to users while protecting data from various security threats. The simulation results confirmed that the proposed PIE system provides higher scalability and stronger security performance than existing blockchain-based medical systems.

Security Analysis

In this section, we will check how the proposed PIE system effectively responds to security threats and analyze whether it is possible to share secure medical data using the proposed PIE system.

Strong Privacy Preservation

When medical data are shared using a network, an external attacker can obtain the medical data through a sniffing or eavesdropping attack. If medical data are leaked, the patient's privacy in the EMR is also exposed. In the proposed PIE system, medical data are encrypted using the dedicated encryption key

for the safe sharing of medical data. As encrypted medical data can only be decrypted by the patient or by a user approved by the patient, the information in the data is not exposed even if the data are stolen. By granting data decryption authority using the PRE technique, the user approved by the patient can decrypt the data using their private key during the data decryption step. The proposed EMR-sharing method prevents the leakage of private information during the EMR-sharing process and ensures safety by eliminating the private key-exchange process for data decryption. If an internal attacker attempts unauthorized access to the patient's information, in our proposed system, *Smart contract*_{RRrequest} verifies the requester's security level and accepts or declines the request depending on the AP set by the patient. The access control scheme using a smart contract can protect the patient's privacy from internal threats.

Data Integrity

The internal attacker can perform forgery attacks by accessing the medical data that medical institutions manage independently. If the original data stored at a medical institution are damaged, it is difficult to recover the data; moreover, it is also significantly challenging to determine whether the data have been forged or altered. These attacks can be effectively prevented by storing and managing EMR-related information such as the hash value of medical data, publicly available medical information, and hospital ID in the blockchain. As the EMR information recorded on the blockchain contains the information at the time of creation, it is easy to check whether the data are damaged. If the data are damaged, they can be quickly restored using the distributed data-sharing method. For an attacker to forge the data stored in the blockchain, they must possess the mighty hash computation power of more than 50% of the entire network and create new blocks faster than other honest nodes propagate them to the network. As meeting the necessary conditions to forge blockchain data is challenging, attackers cannot delete or modify data. Therefore, using a blockchain-based medical system ensures medical data integrity and reliability, thereby enabling safe medical data management and sharing.

Network Security

The external attacker can perform denial-of-service attacks on the PIE system. As a result, the system's operation becomes abnormal and it produces unexpected outputs. The system we propose is directly or indirectly related to patients' lives; therefore, high availability is important. Hence, we use distributed systems such as a medical consortium blockchain and the IPFS. If the attacker breaks down the sharing system, patients cannot share their medical data and physicians or health care providers cannot obtain the required information. However, in the proposed sharing scheme based on blockchain, if the attacker makes a few of the blockchain nodes unavailable, the other nodes can provide the needed services.

Discussion

Principal Findings

The study's principal findings concern implementing integrated management of fragmented EMRs, preventing leakage of personal information of patients during the EMR-sharing

process, and establishing a patient-centered medical data system by granting decryption authority, as outlined in the following list:

1. We designed a blockchain-based PIE medical system that effectively manages and shares medical data. EMRs generated by different medical institutions are managed through a blockchain network to prevent the fragmentation of medical data. Moreover, through the PIE system, duplicate EMRs can be avoided, reducing the cost and wastage of storage space.
2. The PIE system encrypts the patient's medical data and uploads and shares the encrypted EMR and data identification parameters to the network with minimal medical information. Thus, the proposed method fundamentally overcomes the problem of possible leakage of personal data when the data are posted on the blockchain for sharing with other network members. Therefore, privacy preservation required in a system handling sensitive EMR information is guaranteed, making safe EMR management and sharing possible.
3. Our system reinforces the patient's role in the medical system by allowing them to grant decryption rights to their data to other users using re-encryption techniques. If other users (eg, physicians or researchers) wish to use a patient's EMR data, they must obtain a re-encryption key and re-encrypt the EMR data. Building a patient-centered medical data system differs from the existing hospital-centered medical data system in that the patient's role is reinforced in our system.

Limitations and Future Work

Blockchain-based medical systems receive considerable attention as next-generation medical systems that will replace existing medical data management systems, and numerous researchers are conducting various studies. However, blockchain-based medical systems' technological maturity remains at the prototype level. Moreover, as the medical data formats used by different countries or institutions vary considerably, it is challenging to share the medical data. Consequently, research into standardized medical data formats such as the Health Level 7 Fast Healthcare Interoperability Resources [65] is required. To successfully create the next-generation medical environment through a blockchain-based medical system, various and complex issues such as backlash from the medical field, legal ramifications related to medical care, technical limitations, and data standards must be addressed [66,67]. We aim to conduct research on public medical data systems that enable safe sharing of medical data in public networks as well as data security techniques so that EMRs can be used in more diverse fields in the future [68].

Conclusions

This paper presented the PIE system based on a consortium blockchain that allows patients to manage their medical data. The PIE system can securely manage and share EMRs by overcoming the existing blockchain-based medical system's problems. The PIE system uses a distributed data-sharing method and lightweight transaction structure to solve scalability and privacy issues, a chronic problem of blockchain-based

medical systems. By rapidly sharing large-capacity data such as medical images using a distributed data-sharing method, the issues of low processing speed and block sizes of existing blockchains are addressed. Lightweight transactions can store more information in blocks because they contain only minimal information, such as the encrypted EMR metadata and EMR summary information. The vast amount of medical data generated daily is effectively processed and managed using a lightweight transaction structure. The re-encryption-based data encryption method is used to resolve the problem of leakage of data and personal information when sharing EMRs. Even if the EMR encrypted with the dedicated encryption key is leaked

during the sharing process, it cannot be decrypted; therefore, it is safe from the threat of leakage. Honest users wishing to use the patient's data can re-encrypt the EMR by obtaining a re-encryption key from the patient. The EMR-sharing process was performed using smart contracts. Security level-based access control was performed using smart contracts to prevent unauthorized users from using medical data, and re-encryption keys were delivered only to authorized users. As a result, the proposed blockchain-based medical system provides improved security and scalability, enabling efficient and safe medical data sharing.

Acknowledgments

This research was supported by the MSIP (Ministry of Science, Information and Communication Technology, and Future Planning), Korea, under the National Program for Excellence in Software (2018-0-00192) supervised by the IITP (Institute for Information & communications Technology Planning & Evaluation), and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2018R1D1A1B07049043)

Conflicts of Interest

None declared.

Multimedia Appendix 1

Proxy Re-encryption based EMR Encryption & Decryption.
[DOCX File, 242 KB - [jmir_v24i3e29108_app1.docx](#)]

Multimedia Appendix 2

Average data upload time by electronic medical record size.
[PNG File, 49 KB - [jmir_v24i3e29108_app2.png](#)]

Multimedia Appendix 3

Average data upload time to the InterPlanetary File System by electronic medical record.
[PNG File, 52 KB - [jmir_v24i3e29108_app3.png](#)]

Multimedia Appendix 4

Average data download time by electronic medical record size.
[PNG File, 63 KB - [jmir_v24i3e29108_app4.png](#)]

Multimedia Appendix 5

Smart contract execution time for re-encryption key sharing.
[PNG File, 71 KB - [jmir_v24i3e29108_app5.png](#)]

References

1. Parasrampurua S, Henry J. Hospitals' use of electronic health records data, 2015-2017. *ONC Data Brief* (46). 2019. URL: <https://www.healthit.gov/sites/default/files/page/2019-04/AHAEHRUseDataBrief.pdf> [accessed 2022-02-17]
2. 'Non-federal Acute Care Hospital Health IT Adoption and Use,' Health IT Dashboard. Office of the National Coordinator for Health Information Technology. 2019. URL: <http://dashboard.healthit.gov/apps/hospital-health-it-adoption.php> [accessed 2022-02-17]
3. Heart T, Ben-Assuli O, Shabtai I. A review of PHR, EMR and EHR integration: a more personalized healthcare and public health policy. *Health Policy Technol* 2017 Mar;6(1):20-25. [doi: [10.1016/j.hlpt.2016.08.002](https://doi.org/10.1016/j.hlpt.2016.08.002)]
4. Enaizan O, Zaidan AA, Alwi NH, Zaidan BB, Alsalem MA, Albahri OS, et al. Electronic medical record systems: decision support examination framework for individual, security and privacy concerns using multi-perspective analysis. *Health Technol* 2018 Nov 28;10(3):795-822. [doi: [10.1007/s12553-018-0278-7](https://doi.org/10.1007/s12553-018-0278-7)]

5. Yang Z, Huang Y, Jiang Y, Sun Y, Zhang Y, Luo P. Clinical assistant diagnosis for electronic medical record based on convolutional neural network. *Sci Rep* 2018 Apr 20;8(1):6329 [FREE Full text] [doi: [10.1038/s41598-018-24389-w](https://doi.org/10.1038/s41598-018-24389-w)] [Medline: [29679019](https://pubmed.ncbi.nlm.nih.gov/29679019/)]
6. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? *BMJ Glob Health* 2018 Aug 29;3(4):e000798. [doi: [10.1136/bmjgh-2018-000798](https://doi.org/10.1136/bmjgh-2018-000798)] [Medline: [30233828](https://pubmed.ncbi.nlm.nih.gov/30233828/)]
7. Ting DS, Liu Y, Burlina P, Xu X, Bressler NM, Wong TY. AI for medical imaging goes deep. *Nat Med* 2018 May 7;24(5):539-540. [doi: [10.1038/s41591-018-0029-3](https://doi.org/10.1038/s41591-018-0029-3)] [Medline: [29736024](https://pubmed.ncbi.nlm.nih.gov/29736024/)]
8. Savadjiev P, Chong J, Dohan A, Vakalopoulou M, Reinhold C, Paragios N, et al. Demystification of AI-driven medical image interpretation: past, present and future. *Eur Radiol* 2019 Mar 13;29(3):1616-1624. [doi: [10.1007/s00330-018-5674-x](https://doi.org/10.1007/s00330-018-5674-x)] [Medline: [30105410](https://pubmed.ncbi.nlm.nih.gov/30105410/)]
9. Mc Cord KA, Ewald H, Ladanie A, Briel M, Speich B, Bucher HC, RCD for RCTs Initiative and the Making Randomized Trials More Affordable Group. Current use and costs of electronic health records for clinical trial research: a descriptive study. *CMAJ Open* 2019 Feb 03;7(1):23-32 [FREE Full text] [doi: [10.9778/cmajo.20180096](https://doi.org/10.9778/cmajo.20180096)] [Medline: [30718353](https://pubmed.ncbi.nlm.nih.gov/30718353/)]
10. Mc Cord KA, Hemkens LG. Using electronic health records for clinical trials: where do we stand and where can we go? *Can Med Asso J* 2019 Feb 04;191(5):128-133 [FREE Full text] [doi: [10.1503/cmaj.180841](https://doi.org/10.1503/cmaj.180841)] [Medline: [30718337](https://pubmed.ncbi.nlm.nih.gov/30718337/)]
11. Nedungadi P, Jayakumar A, Raman R. Personalized health monitoring system for managing well-being in rural areas. *J Med Syst* 2017 Dec 14;42(1):22. [doi: [10.1007/s10916-017-0854-9](https://doi.org/10.1007/s10916-017-0854-9)] [Medline: [29242996](https://pubmed.ncbi.nlm.nih.gov/29242996/)]
12. Fuse N, Sakurai-Yageta M, Katsuoka F, Danjoh I, Shimizu R, Tamiya G, et al. Establishment of integrated biobank for precision medicine and personalized healthcare: the Tohoku Medical Megabank project. *JMA J* 2019 Sep 04;2(2):113-122 [FREE Full text] [doi: [10.31662/jmaj.2019-0014](https://doi.org/10.31662/jmaj.2019-0014)] [Medline: [33615021](https://pubmed.ncbi.nlm.nih.gov/33615021/)]
13. Mammen JR, Java JJ, Halterman J, Berliant MN, Crowley A, Frey SM, et al. Development and preliminary results of an Electronic Medical Record (EMR)-integrated smartphone telemedicine program to deliver asthma care remotely. *J Telemed Telecare* 2019 Aug 22;27(4):217-230. [doi: [10.1177/1357633x19870025](https://doi.org/10.1177/1357633x19870025)]
14. de Jong MJ, van der Meulen-de Jong AE, Romberg-Camps MJ, Becx MC, Maljaars JP, Cilissen M, et al. Telemedicine for management of inflammatory bowel disease (myIBDcoach): a pragmatic, multicentre, randomised controlled trial. *Lancet* 2017 Sep;390(10098):959-968. [doi: [10.1016/s0140-6736\(17\)31327-2](https://doi.org/10.1016/s0140-6736(17)31327-2)]
15. Helal S, Jain R. Digital health-active and healthy living. *Computer* 2019 Nov;52(11):14-17. [doi: [10.1109/mc.2019.2934590](https://doi.org/10.1109/mc.2019.2934590)]
16. Digital health market - Global industry analysis, size, share, growth, trends and forecast, 2017 - 2025. Cision - PR Newswire. 2017. URL: <https://www.prnewswire.com/news-releases/digital-health-market---global-industry-analysis-size-share-growth-trends-and-forecast-2017---2025-300537263.html> [accessed 2022-02-17]
17. Langer SG, Tellis W, Carr C, Daly M, Erickson BJ, Mendelson D, et al. The RSNA image sharing network. *J Digit Imaging* 2015 Feb 19;28(1):53-61 [FREE Full text] [doi: [10.1007/s10278-014-9714-z](https://doi.org/10.1007/s10278-014-9714-z)] [Medline: [25037586](https://pubmed.ncbi.nlm.nih.gov/25037586/)]
18. Langlotz CP, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, Cook TS, et al. A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/The Academy Workshop. *Radiology* 2019 Jun;291(3):781-791 [FREE Full text] [doi: [10.1148/radiol.2019190613](https://doi.org/10.1148/radiol.2019190613)] [Medline: [30990384](https://pubmed.ncbi.nlm.nih.gov/30990384/)]
19. Edaibat EA, Dever J, Stuban SM. System dynamics simulation modeling of health information exchange (HIE) adoption and policy intervention: a case study in the State of Maryland. *Operat Res Healthc* 2017 Mar;12:60-70. [doi: [10.1016/j.orhc.2017.02.001](https://doi.org/10.1016/j.orhc.2017.02.001)]
20. Yang Y, Li X, Qamar N, Liu P, Ke W, Shen B, et al. Medshare: a novel hybrid cloud for medical resource sharing among autonomous healthcare providers. *IEEE Access* 2018;6:46949-46961. [doi: [10.1109/access.2018.2865535](https://doi.org/10.1109/access.2018.2865535)]
21. Doel T, Shakir DI, Pratt R, Aertsen M, Moggridge J, Bellon E, et al. GIFT-Cloud: a data sharing and collaboration platform for medical imaging research. *Comput Methods Programs Biomed* 2017 Feb;139:181-190 [FREE Full text] [doi: [10.1016/j.cmpb.2016.11.004](https://doi.org/10.1016/j.cmpb.2016.11.004)] [Medline: [28187889](https://pubmed.ncbi.nlm.nih.gov/28187889/)]
22. Jiang T, Chen X, Ma J. Public integrity auditing for shared dynamic cloud data with group user revocation. *IEEE Trans Comput* 2016 Aug 1;65(8):2363-2373. [doi: [10.1109/tc.2015.2389955](https://doi.org/10.1109/tc.2015.2389955)]
23. Ravishankar B, Kulkarni P, Vishnudas M. Blockchain-based database to ensure data integrity in cloud computing environments. In: *Proceedings of the International Conference on Mainstreaming Block Chain Implementation (ICOMBI)*. 2020 Presented at: 2020 International Conference on Mainstreaming Block Chain Implementation (ICOMBI); Feb. 21-22, 2020; Bengaluru, India. [doi: [10.23919/icombi48604.2020.9203500](https://doi.org/10.23919/icombi48604.2020.9203500)]
24. Jain P, Gyanchandani M, Khare N. Differential privacy: its technological prescriptive using big data. *J Big Data* 2018 Apr 13;5(1):15. [doi: [10.1186/s40537-018-0124-9](https://doi.org/10.1186/s40537-018-0124-9)]
25. Fernández-Alemán JL, Señor IC, Lozoya PA, Toval A. Security and privacy in electronic health records: a systematic literature review. *J Biomed Inform* 2013 Jun;46(3):541-562 [FREE Full text] [doi: [10.1016/j.jbi.2012.12.003](https://doi.org/10.1016/j.jbi.2012.12.003)] [Medline: [23305810](https://pubmed.ncbi.nlm.nih.gov/23305810/)]
26. Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. *J Big Data* 2018 Jan 9;5(1):1-18. [doi: [10.1186/s40537-017-0110-7](https://doi.org/10.1186/s40537-017-0110-7)]
27. Roman-Belmonte JM, De LC, Rodriguez-Merchan EC. How blockchain technology can change medicine. *Postgrad Med* 2018 May;130(4):420-427. [doi: [10.1080/00325481.2018.1472996](https://doi.org/10.1080/00325481.2018.1472996)] [Medline: [29727247](https://pubmed.ncbi.nlm.nih.gov/29727247/)]

28. O'Donoghue O, Vazirani AA, Brindley D, Meinert E. Design choices and trade-offs in health care blockchain implementations: systematic review. *J Med Internet Res* 2019 May 10;21(5):e12426 [FREE Full text] [doi: [10.2196/12426](https://doi.org/10.2196/12426)] [Medline: [31094344](https://pubmed.ncbi.nlm.nih.gov/31094344/)]
29. Tai S, Eberhardt J, Klems M. Not ACID, not BASE, but SALT - A transaction processing perspective on blockchains. In: Proceedings of the 7th International Conference on Cloud Computing and Services Science - CLOSER. 2017 Presented at: 7th International Conference on Cloud Computing and Services Science - CLOSER; 2017; Porto, Portugal p. 755-764. [doi: [10.5220/0006408207550764](https://doi.org/10.5220/0006408207550764)]
30. Yaga D, Mell P, Roby N, Scarfone K. Blockchain technology overview. National Institute of Standards and Technology. 2018. URL: <https://nvlpubs.nist.gov/nistpubs/ir/2018/NIST.IR.8202.pdf> [accessed 2022-02-17]
31. Vazirani AA, O'Donoghue O, Brindley D, Meinert E. Implementing blockchains for efficient health care: systematic review. *J Med Internet Res* 2019 Feb 12;21(2):e12439 [FREE Full text] [doi: [10.2196/12439](https://doi.org/10.2196/12439)] [Medline: [30747714](https://pubmed.ncbi.nlm.nih.gov/30747714/)]
32. Szabo N. The idea of smart contracts. Nick Szabo's Papers and Concise Tutorials. 1997. URL: <https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/idea.html> [accessed 2021-06-16]
33. Buterin V. A next-generation smart contract and decentralized application platform. Ethereum White Paper. 2014. URL: https://cryptorating.eu/whitepapers/Ethereum/Ethereum_white_paper.pdf [accessed 2022-02-17]
34. Wood G. Ethereum: a secure decentralised generalised transaction ledger. Ethereum Yellow Paper. 2022. URL: <https://ethereum.github.io/yellowpaper/paper.pdf> [accessed 2022-02-17]
35. Griggs KN, Ossipova O, Kohlios CP, Baccarini AN, Howson EA, Hayajneh T. Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. *J Med Syst* 2018 Jun 06;42(7):130. [doi: [10.1007/s10916-018-0982-x](https://doi.org/10.1007/s10916-018-0982-x)] [Medline: [29876661](https://pubmed.ncbi.nlm.nih.gov/29876661/)]
36. Jamil F, Hang L, Kim K, Kim D. A novel medical blockchain model for drug supply chain integrity management in a smart hospital. *Electronics* 2019 May 07;8(5):505. [doi: [10.3390/electronics8050505](https://doi.org/10.3390/electronics8050505)]
37. Khatoun A. A blockchain-based smart contract system for healthcare management. *Electronics* 2020 Jan 03;9(1):94. [doi: [10.3390/electronics9010094](https://doi.org/10.3390/electronics9010094)]
38. Wüst K, Gervais A. Do you need a blockchain? In: Proceedings of the Crypto Valley Conference on Blockchain Technology (CVCBT). 2018 Presented at: Crypto Valley Conference on Blockchain Technology (CVCBT); June 20-22, 2018; Zug, Switzerland. [doi: [10.1109/cvcbt.2018.00011](https://doi.org/10.1109/cvcbt.2018.00011)]
39. Yang G, Li C, Marstein KE. A blockchain - based architecture for securing electronic health record systems. *Concurrency Computat Pract Exper* 2019 Aug 04;33(14):e5479. [doi: [10.1002/cpe.5479](https://doi.org/10.1002/cpe.5479)]
40. Hau YS, Lee JM, Park J, Chang MC. Attitudes toward blockchain technology in managing medical information: survey study. *J Med Internet Res* 2019 Dec 09;21(12):e15870 [FREE Full text] [doi: [10.2196/15870](https://doi.org/10.2196/15870)] [Medline: [31815676](https://pubmed.ncbi.nlm.nih.gov/31815676/)]
41. Mettler M. Blockchain technology in healthcare: the revolution starts here. In: Proceedings of the IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom). 2016 Presented at: IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom); Sept. 14-16, 2016; Munich, Germany. [doi: [10.1109/healthcom.2016.7749510](https://doi.org/10.1109/healthcom.2016.7749510)]
42. Blockchain: the chain of trust and its potential to transform healthcare – our point of view. Office of the National Coordinator for Health Information Technology. 2016. URL: https://www.healthit.gov/sites/default/files/8-31-blockchain-ibm_ideation_challenge_aug8.pdf [accessed 2022-02-17]
43. Azaria A, Ekblaw A, Vieira T, Lippman A. MedRec: using blockchain for medical data access and permission management. In: Proceedings of the 2nd International Conference on Open and Big Data (OBD). 2016 Presented at: 2nd International Conference on Open and Big Data (OBD); Aug. 22-24, 2016; Vienna, Austria. [doi: [10.1109/obd.2016.11](https://doi.org/10.1109/obd.2016.11)]
44. Uddin MA, Stranieri A, Gondal I, Balasubramanian V. Continuous patient monitoring with a patient centric agent: a block architecture. *IEEE Access* 2018;6:32700-32726. [doi: [10.1109/access.2018.2846779](https://doi.org/10.1109/access.2018.2846779)]
45. Maslove DM, Klein J, Brohman K, Martin P. Using blockchain technology to manage clinical trials data: a proof-of-concept study. *JMIR Med Inform* 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: [10.2196/11949](https://doi.org/10.2196/11949)] [Medline: [30578196](https://pubmed.ncbi.nlm.nih.gov/30578196/)]
46. Guo R, Shi H, Zheng D, Jing C, Zhuang C, Wang Z. Flexible and efficient blockchain-based abe scheme with multi-authority for medical on demand in telemedicine system. *IEEE Access* 2019;7:88012-88025. [doi: [10.1109/access.2019.2925625](https://doi.org/10.1109/access.2019.2925625)]
47. Hylock RH, Zeng X. A blockchain framework for patient-centered health records and exchange (HealthChain): evaluation and proof-of-concept study. *J Med Internet Res* 2019 Aug 31;21(8):e13592 [FREE Full text] [doi: [10.2196/13592](https://doi.org/10.2196/13592)] [Medline: [31471959](https://pubmed.ncbi.nlm.nih.gov/31471959/)]
48. Wu S, Du J. Electronic medical record security sharing model based on blockchain. In: Proceedings of the 3rd International Conference on Cryptography, Security and Privacy. 2019 Presented at: ICCSP 2019: 2019 the 3rd International Conference on Cryptography, Security and Privacy; January 19 - 21, 2019; Kuala Lumpur Malaysia p. 13-17. [doi: [10.1145/3309074.3309079](https://doi.org/10.1145/3309074.3309079)]
49. Abdellatif AA, Al-Marridi AZ, Mohamed A, Erbad A, Chiasserini CF, Refaey A. ssHealth: toward secure, blockchain-enabled healthcare systems. *IEEE Network* 2020 Jul;34(4):312-319. [doi: [10.1109/mnet.011.1900553](https://doi.org/10.1109/mnet.011.1900553)]
50. Shao J, Liu P, Wei G, Ling Y. Anonymous proxy re-encryption. *Security Comm Networks* 2011 May 04;5(5):439-449. [doi: [10.1002/sec.326](https://doi.org/10.1002/sec.326)]

51. Polyakov Y, Rohloff K, Sahu G, Vaikuntanathan V. Fast proxy re-encryption for publish/subscribe systems. *ACM Trans Priv Secur* 2017 Oct 26;20(4):1-31. [doi: [10.1145/3128607](https://doi.org/10.1145/3128607)]
52. Kim S, Lee I. IoT device security based on proxy re-encryption. *J Ambient Intell Human Comput* 2017 Nov 30;9(4):1267-1273. [doi: [10.1007/s12652-017-0602-5](https://doi.org/10.1007/s12652-017-0602-5)]
53. Vijayakumar V, Priyan MK, Ushadevi G, Varatharajan R, Manogaran G, Tarare PV. E-Health cloud security using timing enabled proxy re-encryption. *Mobile Netw Appl* 2018 May 21;24(3):1034-1045. [doi: [10.1007/s11036-018-1060-9](https://doi.org/10.1007/s11036-018-1060-9)]
54. Manzoor A, Liyanage M, Braeke A, Kanhere S, Ylianttila M. Blockchain based proxy re-encryption scheme for secure IoT data sharing. In: *Proceedings of the IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. 2019 Presented at: IEEE International Conference on Blockchain and Cryptocurrency (ICBC); May 14-17, 2019; Seoul, Korea (South). [doi: [10.1109/bloc.2019.8751336](https://doi.org/10.1109/bloc.2019.8751336)]
55. Benet J. Ipfs-content addressed, versioned, p2p file system. *arXiv* 2014:1-11 [FREE Full text]
56. DeSalvo K, Galvez E. Connecting health and care for the nation: a shared nationwide interoperability roadmap – version 1.0. Office of the National Coordinator for Health Information Technology. 2015. URL: <https://tinyurl.com/yhnjbsw4> [accessed 2021-06-16]
57. Mackey TK, Kuo T, Gummadi B, Clauson KA, Church G, Grishin D, et al. ‘Fit-for-purpose?’ – challenges and opportunities for applications of blockchain technology in the future of healthcare. *BMC Med* 2019 Mar 27;17(1):1-17. [doi: [10.1186/s12916-019-1296-7](https://doi.org/10.1186/s12916-019-1296-7)] [Medline: [30914045](https://pubmed.ncbi.nlm.nih.gov/30914045/)]
58. Greenhalgh T, Hinder S, Stramer K, Bratan T, Russell J. Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. *Br Med J* 2010 Nov 16;341(nov16 1):c5814 [FREE Full text] [doi: [10.1136/bmj.c5814](https://doi.org/10.1136/bmj.c5814)] [Medline: [21081595](https://pubmed.ncbi.nlm.nih.gov/21081595/)]
59. de Oliveira MT, Michalas A, Groot A, Marquering H, Olabariaga S. Red Alert: break-glass protocol to access encrypted medical records in the cloud. In: *Proceedings of the IEEE International Conference on E-health Networking, Application & Services (HealthCom)*. 2019 Presented at: IEEE International Conference on E-health Networking, Application & Services (HealthCom); Oct. 14-16, 2019; Bogota, Colombia. [doi: [10.1109/healthcom46333.2019.9009598](https://doi.org/10.1109/healthcom46333.2019.9009598)]
60. Harron K, Dibben C, Boyd J, Hjern A, Azimae M, Barreto ML, et al. Challenges in administrative data linkage for research. *Big Data Soc* 2017 Dec 05;4(2):2053951717745678 [FREE Full text] [doi: [10.1177/2053951717745678](https://doi.org/10.1177/2053951717745678)] [Medline: [30381794](https://pubmed.ncbi.nlm.nih.gov/30381794/)]
61. Er-Rajy L, El KM, El GM, Achbarou O. Blockchain: bitcoin wallet cryptography security, challenges and countermeasures. *J Internet Bank Comm* 2017;22(3):1-29 [FREE Full text]
62. Blackshear S, Chalkias K, Chatzigiannis P, Faizullahoy R, Khaburzaniya I, Kokoris-Kogias E, et al. Reactive key-loss protection in blockchains. In: *Financial Cryptography and Data Security*. Berlin, Heidelberg: Springer; 2021:431-450.
63. Cachin C. Architecture of the hyperledger blockchain fabric. IBM Research - Zurich. 2016. URL: https://www.zurich.ibm.com/dcc/papers/cachin_dccl.pdf [accessed 2022-02-18]
64. Apache CouchDB. 2019. URL: <http://couchdb.apache.org/> [accessed 2022-02-17]
65. Bender D, Sartipi K. HL7 FHIR: an agile and RESTful approach to healthcare information exchange. In: *Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems*. 2013 Presented at: IEEE International Symposium on Computer-Based Medical Systems; June 20-22, 2013; Porto, Portugal. [doi: [10.1109/cbms.2013.6627810](https://doi.org/10.1109/cbms.2013.6627810)]
66. CFR - Code of Federal Regulations Title 21. Part 50: protection of human subjects. Department of Health and Human Services - FDA. 2017. URL: <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcr/CFRSearch.cfm?CFRPart=50> [accessed 2021-06-16]
67. The HIPAA privacy rule. U.S. Department of Health and Human Services. 2017. URL: <http://www.hhs.gov/hipaa/> [accessed 2021-06-16]
68. Esposito C, De Santis A, Tortora G, Chang H, Choo KR. Blockchain: a panacea for healthcare cloud-based data security and privacy? *IEEE Cloud Comput* 2018 Jan;5(1):31-37. [doi: [10.1109/MCC.2018.011791712](https://doi.org/10.1109/MCC.2018.011791712)]

Abbreviations

AP: access policy

CA: certificate authority

EMR: electronic medical record

IPFS: InterPlanetary File System

ONC: Office of the National Coordinator for Health Information Technology

PIE: patient information exchange

PRE: proxy re-encryption

TTP: trusted third party

Edited by A Mavragani; submitted 26.03.21; peer-reviewed by M Platt, L Wang, S Wilson; comments to author 11.05.21; revised version received 05.07.21; accepted 17.01.22; published 22.03.22.

Please cite as:

Lee S, Kim J, Kwon Y, Kim T, Cho S

Privacy Preservation in Patient Information Exchange Systems Based on Blockchain: System Design Study

J Med Internet Res 2022;24(3):e29108

URL: <https://www.jmir.org/2022/3/e29108>

doi: [10.2196/29108](https://doi.org/10.2196/29108)

PMID: [35315778](https://pubmed.ncbi.nlm.nih.gov/35315778/)

©Sejong Lee, Jaehyeon Kim, Yongseok Kwon, Teasung Kim, Sunghyun Cho. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 22.03.2022. 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.

Original Paper

Improving Diagnosis Through Digital Pathology: Proof-of-Concept Implementation Using Smart Contracts and Decentralized File Storage

Hemang Subramanian¹, BE, MS, EMBA, PhD; Susmitha Subramanian², MBBS, DCP, DNB

¹Department of Information Systems and Business Analytics, Florida International University, Miami, FL, United States

²Department of Clinical Pathology, Apollo Super Speciality Hospitals, Bangalore, India

Corresponding Author:

Hemang Subramanian, BE, MS, EMBA, PhD

Department of Information Systems and Business Analytics

Florida International University

11200 NW 8th street

Miami, FL, 33193

United States

Phone: 1 3053482830

Fax: 1 3053482800

Email: hsubrama@fiu.edu

Abstract

Background: Recent advancements in digital pathology resulting from advances in imaging and digitization have increased the convenience and usability of pathology for disease diagnosis, especially in oncology, urology, and gastroenteric diagnosis. However, despite the possibilities to include low-cost diagnosis and viable telemedicine, digital pathology is not yet accessible owing to expensive storage, data security requirements, and network bandwidth limitations to transfer high-resolution images and associated data. The increase in storage, transmission, and security complexity concerning data collection and diagnosis makes it even more challenging to use artificial intelligence algorithms for machine-assisted disease diagnosis. We designed and prototyped a digital pathology system that uses blockchain-based smart contracts using the nonfungible token (NFT) standard and the Interplanetary File System for data storage. Our design remediates shortcomings in the existing digital pathology systems infrastructure, which is centralized. The proposed design is extendable to other fields of medicine that require high-fidelity image and data storage. Our solution is implemented in data systems that can improve access quality of care and reduce the cost of access to specialized pathological diagnosis, reducing cycle times for diagnosis.

Objective: The main objectives of this study are to highlight the issues in digital pathology and suggest that a software architecture-based blockchain and the Interplanetary File System create a low-cost data storage and transmission technology.

Methods: We used the design science research method consisting of 6 stages to inform our design overall. We innovated over existing public-private designs for blockchains but using a 2-layered approach that separates actual file storage from metadata and data persistence.

Results: Here, we identified key challenges to adopting digital pathology, including challenges concerning long-term storage and the transmission of information. Next, using accepted frameworks in NFT-based intelligent contracts and recent innovations in distributed secure storage, we proposed a decentralized, secure, and privacy-preserving digital pathology system. Our design and prototype implementation using Solidity, web3.js, Ethereum, and node.js helped us address several challenges facing digital pathology. We demonstrated how our solution, which combines NFT smart contract standard with persistent decentralized file storage, solves most of the challenges of digital pathology and sets the stage for reducing costs and improving patient care and speed of diagnosis.

Conclusions: We identified technical limitations that increase costs and reduce the mass adoption of digital pathology. We presented several design innovations using NFT decentralized storage standards to prototype a system. We also presented the implementation details of a unique security architecture for a digital pathology system. We illustrated how this design can overcome privacy, security, network-based storage, and data transmission limitations. We illustrated how improving these factors sets the stage for improving data quality and standardized application of machine learning and artificial intelligence to such data.

KEYWORDS

digital pathology; nonfungible token standard; decentralized storage; security and patient data confidentiality using design; pathology; storage; security; confidentiality; data; design; diagnosis; proof of concept; implementation; software; blockchain; limitation; privacy

Introduction

Background

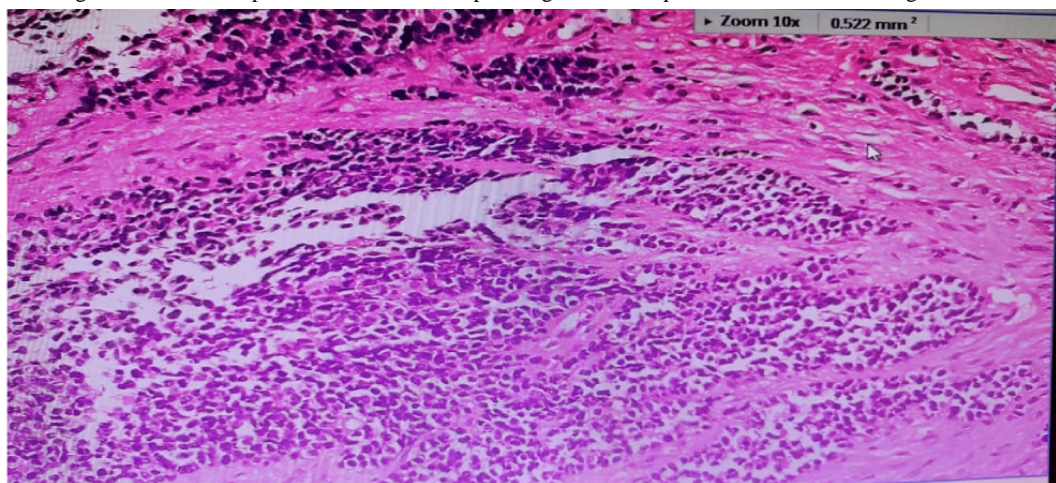
Digital pathology, a subfield of pathology, deals with information management enabled by digital technologies such as high-resolution imaging, computer storage, and network connectivity [1,2]. In digital pathology, the information generated from digitizing slides containing microscopic images of biological samples is shared over a computer network, with the expectation of faster diagnosis and remote access to medical experts. Digital pathology has several advantages over conventional pathology, which bases diagnosis on glass slides and microscopic examination of images on slides [3]. In digital pathology, extremely high-resolution images enable physicians to tag, expand, share, and analyze specific sections of the image slides to diagnose a disease. More recently, the deployment of deep learning algorithms to recognize patterns in high-resolution digital images has been shown to reduce the time taken to diagnose particular diseases [3,4]. Digital pathology also allows pathologists and referring physicians to share data with other experts in clinical diagnostics to concur concerning diagnosis and treatment pathways for specific diseases. As an example, trained superspecialists in oncological pathology, such as the coauthor (SS) of this paper, can provide a very detailed clinical diagnosis of the specific type of cancer, the rate of spread, the speed of spread, the stage, and the types of cells and tissues affected by detailed examination of slides.

Furthermore, the slides and diagnosis can be shared with other senior experts to confirm the diagnosis. These specialists can further suggest and approve treatments for the patients, thus reducing cycle time and improving diagnosis accuracy through

telediagnosis [5]. Thus, digital pathology can provide an accurate, timely, and concise diagnosis of the disease while providing the ability to share such diagnoses with other specialists.

Digital pathology uses technology that includes high-resolution remote and internet-based microscopy and high-resolution scans of glass slides, which can have multiple advantages. First, digital pathology benefits patients in locations where access to medical care is challenging and where the number of clinicians, physicians, qualified pathologists, and hospital systems are scarce [6]. Second, digitizing images at high resolution provides physicians and specialists with ease of use, thereby reducing the use of a microscope. However, they can view high-resolution images on a computer screen, thus reducing fatigue to the eye [1,7]. Third, physicians and doctors can focus on specific areas of the digital image by using input-output devices such as keyboards, mice, and digital pencils to write notes on specific areas of the images. Figure 1 shows the images and magnified sections of imaged slides that pathologists use to diagnose the medical condition. Fourth, physicians can share such data with other experts or senior practitioners in the field, who can later arrive at a consensus concerning the diagnosis about the progress of disease and recommend treatment pathways for patients. Owing to these advantages of personal ease of use and shareability, the role of digital pathology in providing timely, accurate, and concise prognosis and pathways to clinical recovery has been recognized by the medical fraternity. Hospital systems are slowly transitioning to a fully digital diagnosis mode, reducing physical storage requirements for glass slides and partially automating the transmission, attainment of agreement among experts, and treatment pathway recommendations for patients [5].

Figure 1. An illustration of a digital pathology image from a slide used in an actual diagnosis by the coauthor (SS). The image is from a high-resolution microscopic scan and digitized slide that is presented to the clinical pathologist on a computer screen for further diagnosis.



Infrastructure Challenges for Digital Pathology

However, there are several challenges with the adoption of digital pathology to make it widely accessible. These challenges relate to the transmission, storage, retrieval, and uniform clinical diagnosis using algorithms from slide or slides and image or images derived from tissue and biosample scans. Digital image scans are made possible by remote and internet-based microscopy or through a high-resolution scan of microscopic slides into images of much higher resolution (100,000 dpi to 1,000,000 dpi). Such scanned images also require high-resolution storage, for the data not to be compressed. While transmitting such high-resolution images, transmission could lead to data loss if images are compressed and decompressed, potentially affecting the consensus for diagnosis across the network. Compression and decompression of data lead to loss of information, which can potentially lead to an inaccurate diagnosis or inaccurate data training from these samples [8]. In addition, inaccurate images and matched diagnoses create flawed training data sets, reducing the prediction accuracy when deep learning models are applied to such data sets. Although network-attached storage, cloud-based storage, commercial databases, and workflow systems such as enterprise resource planning frameworks are plausible solutions to accomplish security and transmissibility of related data (ie, image scans or diagnoses), these centralized software systems are expensive to own and increase the total cost of ownership for pathological systems, which makes disease diagnoses via pathology unaffordable [6].

In centralized systems, the cloud provider or database network administrator controls the security of the network and patient data. Such centralized data control could often violate patient privacy and be susceptible to security breaches, ransomware attacks, and other data security problems. As a result, both the storage and network-based transmission of such high-resolution images are challenging. Second, obtaining the necessary audit or data transmission logs of such algorithms is difficult in centralized systems, which are database-centric and often need to interface between different hospital and clinical systems managed by different entities [9,10]. Third, although most research in the digital pathology area has automated deep learning techniques for pattern recognition with images, such application of algorithms will only fructify when a sufficient corpus of images combined with human-expert classification or training is present. Specialized diagnosis required to detect rare diseases such as sickle cell disease or certain types of cancer such as bone marrow cancer makes it increasingly challenging to obtain large training samples of digital images in high-resolution biosamples with human classification. A decentralized, secure, and standardized storage of high-resolution images and appropriate expert classification of such data that includes specific sections of tissue scans, physically marked and noted with the clinical disease diagnosis, can potentially solve this problem. Such uniform storage of images, diagnoses, and associated metadata information such as sections of the images, which help physicians diagnose the disease, is essential. Such data can lead to the creation of a proper training data set for deep learning, which can enhance the capabilities of automated disease diagnosis in the future.

The fields of artificial intelligence and machine learning, such as deep learning, are applied to remote and internet-based microscopy-derived images, using high-quality training samples, wherein the trained models can aid in faster diagnosis [1,3,4,11]. Such machine-assisted diagnoses can aid physicians by complementing their knowledge through pattern detection.

This paper highlights the challenges in digital pathology associated with secure data storage, secure data transmission, data privacy, and storage and transmission of high-resolution images and associated diagnosis from physicians. Furthermore, we propose a unique design with smart contracts and the Ethereum blockchain, which uses the Ethereum Request for Comments (ERC)-721 standard to store and transmit images while tracking ownership [12]. We propose an innovative decentralized storage and encryption scheme separate from the public blockchain [13]. We also document how these challenges can be overcome. We outline why and how such a design using smart contracts and blockchains can overcome issues in training accuracy using machine learning and artificial intelligence techniques for pattern detection from slides with a higher image resolution.

Advantages of Digital Pathology

The primary responsibility of practicing pathologists is to maintain the confidentiality of patient data and diagnoses. In digital pathology, it is a typical pattern that multiple pathologists and diagnostic programs have access to the data uploaded from one of the following sources: (1) high-resolution tissue photography, (2) scanned slides that constitute images from microscopes, or (3) remote and internet-based microscopes that image human tissue samples and store and transmit these images. The patient's medical record and diagnosis should be shared with those with the *right to know*. Data stored are expected to adhere to laws such as Health Insurance Portability and Accountability Act and General Data Protection Regulation. Similarly, only the pathologist who accesses the digital pathology platform should view these high-resolution images, make modifications, and make a diagnosis. Although pathologists and other specialists can access these images and slides on a case-by-case basis, the primary owner of the data is the patient. The patient can choose to remove access to the physician, or the hospital system might invalidate the data access after a legally mandated time frame (eg, 3 years).

In addition, these images should be stored only for an extended period, as mandated by law, and retrieved cost-effectively for purposes such as reference [5]. As discussed in the previous section, these high-resolution images will require significant storage and secure high-bandwidth transmission to transfer images, diagnoses, and audit logs. If these conditions are not met, digital pathology costs will become financially unviable to adopt and use for diagnosis. In the long term, storage requirements and cost reduction are essential for pathologists to ensure data accuracy, privacy, and access control [9,10]. Long-term storage costs, the transmission of high-resolution images about digital pathology to different owners on the network, and the high costs associated with the security, privacy, and auditability of such data create 3 main issues for the success

of digital pathology as an effective practice. The challenges in digital pathology are summarized in [Figure 2](#).

In digital pathology, the image resolution of slides containing images of tissues or other human body samples such as blood vessels is comparatively high. It often ranges between 1× and 400× of slides obtained using traditional microscopy in nondigital pathology [14]. The advantage of such platforms is that multiple pathologists can simultaneously access the slide if they have the corresponding software and authorized access to the slides. In addition, pathologists can share independent diagnoses on the web after analyzing these images. As a result, significant time is saved in prognosis or diagnosis and the treatment compared with scenarios where microscopic glass slides would have to be transported physically. A second or third opinion must be obtained to draw a fail-safe conclusion if the severity of the disease is high, in which case reports, images, and other patient data must be shared by maintaining the privacy and patient confidentiality protocols.

Artificial intelligence algorithms can aid physicians in detecting patterns from images, especially in pathology, where cancer growth creates patterns in body tissues. Deep learning models can be trained with sufficient samples to detect and code sections of images with a specific diagnosis [3,11]. Such models can recognize patterns of spread, the spread stage, and the spread level in cancerous cells after appropriate training from a large corpus of samples that physicians have accurately classified. In addition, prognostic factors that can aid in treatment can be determined. Although storing and documenting preliminary case data for human diagnosis is equally important, such data can later be used downstream as inputs to training and classification algorithms to automate the detection and diagnosis of pathological conditions. The use of deep learning algorithms

such as convolutional neural networks and recurrent neural networks combined with support from graphical processing units offered by NVIDIA and AMD provides significant strength to such automated analysis [3,11]. Often, such trained algorithms have lower accuracy than required because of the uniqueness of individual cases. If irregular patterns do not fit into the diagnosis, then such cases have to be manually studied, and such cases could be labeled independently as a rare possibility.

The goal of such a digital pathology could evolve to enable automatic diagnosis through the accrual of a proper training data set and human classification algorithms. When such data sets become available and deep learning algorithms are trained appropriately on human-classified data, the time to diagnosis will naturally decrease. In addition, data from digital pathology can be used for multiple purposes such as teaching, research, and continuing medical education seminars. Here, data retrieval is fast; access is provided on a distributed platform that caters to both privacy and security, and search and retrieval and display of data in the digital formats used in traditional pathology is proper. [Figure 3](#) shows a sequence diagram that captures a typical pathology workflow for diagnosing a patient’s condition using digital pathology.

Pathologists can choose to expand and focus on certain sections of the slides at a higher resolution to diagnose the disease involved and create notes on these slides by annotating sections of the scanned images. Later, multiple pathologists or other specialists can view the same slides to create a diagnosis and discuss treatment pathways. These modified slides with annotations from experts are stored back in the centralized system for retrieval and use, in other cases, for research or teaching purposes.

Figure 2. Technical challenges in digital pathology.

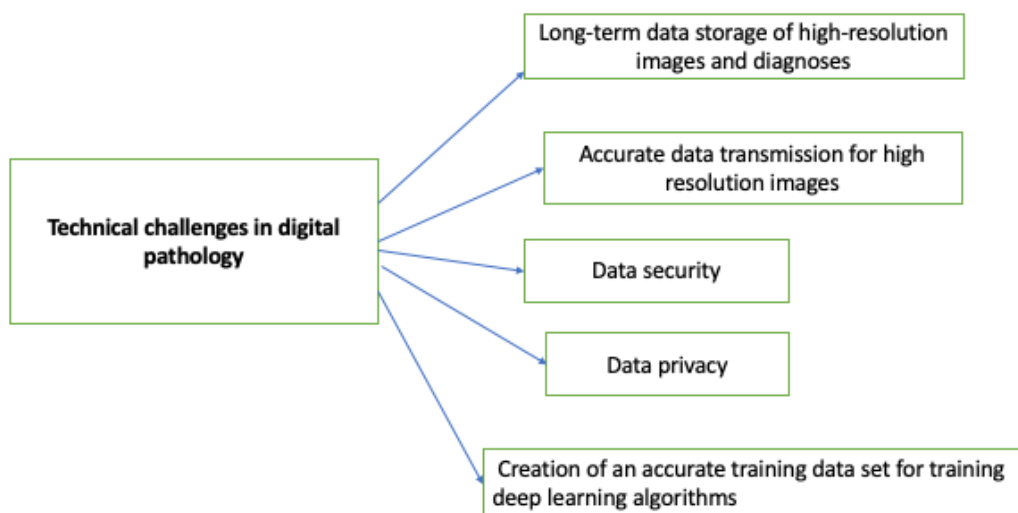


Figure 3. Workflow of digital pathology system.

Blockchains, Nonfungible Tokens, and Decentralized Storage (Interplanetary File System) Applied to Health Care

Blockchains address the challenges of system interoperability, security, privacy, nonrepudiation, storage persistence, decentralized storage, decentralized validation of transactions, and accessibility for all data [13]. Before the advent of blockchains, anonymized and privacy-controlled single point of access for multiple data sources was provided for each user. Similarly, assuring privacy and user-controlled data and tracing or auditing data transmission records was difficult because the data resided in private centrally controlled repositories. Such data repositories are administered by a central administrator, who is often the single weak link for ensuring data security and privacy. In medical record storage systems, owing to multiple disparate administration end points involving different organizations, data are shared among multiple administrators, creating a weak link for a privacy violation. Because blockchains remove the need to centralize data storage and control data flow and modifications, they have several applications in health care [15]. Blockchains have been shown to provide a cost-effective and robust decentralized technology for a variety of health care applications, such as tracking clinical trials [16], enabling tamper resistance [17] for mobile health data, removing trust deficiencies during the COVID-19 pandemic [18], and eliminating substandard drugs from the supply chain [19].

Blockchains have evolved to support programming logic by including variables and loop structures using high-level programming languages such as Solidity [20] and a remote and internet-based machine that interprets and executes the compiled programs. These programs are compiled and deployed onto remote and internet-based machines that validate transactional and semantic logic in programs. Remote and internet-based machines understand opcodes (or machine-interpretable code), which are outputs of a compilation of a higher-level programming language such as Solidity. Examples of blockchains include Bitcoin (which supports a rudimentary scripting language), Ethereum, Cardano, and Solana. These smart contracts enable web- or mobile-based decentralized applications that interact with the underlying blockchain and remote and internet-based machine based on user input.

Although blockchains such as Hyperledger, Ethereum, and Cardano have been shown to support data storage and programmability of contracts, it remains challenging to ensure that such data storage can happen cost-effectively on the blockchain (on-chain storage). The governance structures of public blockchains require application writers and data creators to pay a transaction fee, which is often proportional to the bandwidth and space used on the blockchain.

This is because storing even small amounts of binary data on a public or private blockchain, that is, on-chain, as a transaction record would cost several hundred dollars. Although several papers in the medical informatics field have chosen this costly approach to store data on the blockchain (or have used a Hyperledger framework with a private blockchain or a consortium blockchain to store data), such storage mechanisms are inefficient for digital pathology, which requires more extensive storage and other challenges, as shown in Figure 2.

This is because image storage requires large quantities of cost-effective data storage. Image transmission on the network with such large images is nearly impossible, given today's email client and other networking protocols, such as Simple Mail Transfer Protocol. The only possible likelihood is to store images from a local repository that distributes the image onto a network file storage and makes the file addressable.

Another challenge with such an approach is that the data stored in the blockchain are public information and can be effectively downloaded from the remote and internet-based machine by any user as it is stored. Therefore, unless the stored data are encrypted in an unidentifiable way using the user identifiers of both the sender and the recipients, such data are of no use. In this study, we propose a 2-layer blockchain solution for nonfungible tokens using the ERC-721 standard for nonfungible tokens (NFTs) that separate metadata from file storage and promises authenticity, security, distributed storage, and ease of use [21]. We use this 2-level architecture that provides affordability for use cases such as digital pathology, which requires extremely high-resolution and high-density storage. In addition, we use an existing well-used standard supported by sizable open-source marketplaces in the context of digital art to support all use cases necessary to create a cost-effective digital pathology system. Such a system will provide traceability

of each transaction by using transaction records and logs and cost-effectively enable storage and transmission. We propose application-level encryption of metadata URI's stored on-chain and associated with the nonfungible token. We prototype a system and demonstrate how such a decentralized workflow for storage, retrieval, and transmission occurs while catering to various design requirements for such a system.

Decentralized Data Storage Using the Interplanetary File System

Although blockchain technology provides several advantages noted above and health care can be shared efficiently among peers (hospitals or physicians) in a health care system, there has never been a justification to store such health care data on-chain [22]. Particularly in digital pathology, where high-resolution image scans require significant storage, the feasibility of storing data on-chain is impossible because of the infrastructure costs involved in storing, validating, and retrieving such data. Particularly for mass-adoption blockchains such as Ethereum, Solana, and ADA, data storage and transaction fees would significantly exceed the affordability of the digital service compared with the traditional mechanism of glass slides. A simple estimate for the Ethereum blockchain stated that it costs between US \$17,000 and US \$76,000 to store 1 MB of data on-chain. In addition, the blockchain does not provide version control or other features to optimize data storage such as the Interplanetary File System (IPFS). The operation of the IPFS for file storage and retrieval is described as follows.

When a file is added to the IPFS, it is split into smaller chunks, cryptographically hashed, and given a unique fingerprint called the content identifier (CID). The CID acts as a permanent record of the file as it currently exists. When other nodes (or a program using JavaScript Object Notation and Remote Procedure Call) look up the file using the unique CID, they ask their peer nodes about storing the content referenced by the file's CID. The IPFS provides a mechanism to identify (pin) records to hosts on the network. If the content is not pinned, it can be deleted to access the more recently accessed files. Versioning of files is maintained by creating a new CID because of which files stored on the IPFS are resistant to tampering and censorship on the network. The IPFS uses a distributed hash table to identify addresses and images. The IPFS facilitates the removal of duplicate files and enables the creation of version-control history [2]. The IPFS optimizes the storage and retrieval of frequently accessed files through a cache mechanism, where frequently accessed files are prioritized using hash tables that rank files based on their access frequencies. On the IPFS, each file can be individually accessed with its hash address from the internet; the IPFS is a content-addressed block storage system with features such as high throughput, security with hash mapping of transactions, and concurrent access of files.

NFT Standard for Digital Pathology

The NFT standard is a unique and noninterchangeable unit of data (or a data pointer) stored on the blockchain. Although other token standards such as the cryptocurrency token standard ERC-20 have been popular, the NFT standard authenticates and associates reproducible digital files such as photos, videos, and audio on the Ethereum blockchain. The ERC-20 standard is

used to implement the cryptotoken functionality. It is usually used to create a token for sourcing funding during an initial coin offer and other such uses, enabling token holders to participate in blockchain governance. The major feature of ERC-20 is that it is tradable against any other ETH-compatible token, on crypto exchanges, or coin-swapping liquidity provider platforms such as Uniswap. As a result, we have to use the ERC-721 or NFT standard to implement the image storage, authentication, ownership, and digital rights in the context of digital pathology. NFTs use a digital ledger to provide a public certificate of authenticity or proof of ownership. The NFT standard has wide adoption for various multimedia-based token creation and has seen significant adoption among the art collectors. Similarly, more recent innovations in virtual reality, augmented reality, and the metaverse have enabled the creation and sale of digital assets using this standard.

Overall, if we store the characteristics of the image off-chain on the IPFS (eg, in a metadata file) and have the metadata's unique CID referenced on the blockchain and then minted into an NFT, then it is possible to separate token ownership from token access. The ownership is documented in the metadata file, whereas the access is implemented via functions supported by the ERC-721 standard. This architecture provides us with the flexibility to enable the separation of ownership of the token from the access properties. It also enables us to securely transfer access (or replicate access) after signing the token by the owner, enables token burn, and assigns access to others. In addition, such a design enables the process of minting, replication, and burning of the token. The NFT standard uses these two technologies, that is, the NFT standard smart contract on the Ethereum blockchain and the IPFS for decentralized file storage.

The Goal of This Study

In this paper, we analyze the following two research questions (RQs) using the design science research methodology:

RQ1: How does blockchain-enabled digital pathology improve data sanctity, privacy, data validity, and cost-effective, secure network storage?

RQ2: Does such a system design improve physician diagnosis by reducing the cost of storage, improving security diagnosis times, and improving machine learning capabilities by improving the accuracy of data stored?

Methods

Overview

We use the design science research process (DSRP) methodology used in information systems and computer science disciplines and expert recommendations. Our solution prototype proposes an innovative smart contract-based digital pathology system using public infrastructure to simplify high-resolution image storage and provide a decentralized, secure, and privacy-preserving data-sharing solution. Figure 4 shows the research methodology [23].

DSRP is a well-established research methodology in information systems, computer science, and information technology management. It is used mainly to create design-based system

solutions for practical problems, for which there are no previous solutions. As the name states, DSRP applies design science to create a software- or system-based prototype that addresses the research problem. Using the prototype, researchers validate the assumptions they made. We use the following steps in DSRP to implement and test our prototype: First, we describe the problem and define its importance. The challenges in digital pathology are described in the *Introduction* section. We obtain the problem definition (challenges) from practicing pathologists—the coauthor (SS) of this paper—through a formal procedure of requirement gathering. We share a questionnaire about using the digital pathology system and record how they faced challenges in day-to-day work. Second, the objectives of the solution are clarified in terms of the benefits of improving the quality of care and maintaining accurate data for creating clean data sets. These questions are documented in [Multimedia Appendix 1](#). [Table 1](#) defines the key problems that we aim to address using our solution and the main objectives.

In step 3, we design and implement a proof-of-concept use principles of systems analysis and design to create a model-view-controller architecture-based system with JavaScript, web3.js, node.js, and Ethereum stack combined with the IPFS for long-term data storage. The following section describes the creation of the NFT-based IPFS that meets the various objectives of the solution. We later discuss the confidentiality requirements and application-level encryption of metadata files for storage. In the fourth step, we demonstrate the artifact and the solution prototype. Fifth, we evaluate the artifact for whether it indeed met all the design objectives for the problem it indeed set out to solve. Finally, we document the results of our experiments and describe how our solution benefits the wider digital pathology community and the contributions of this study. We also believe that our solution can be appropriated for other high-density storage requirements in the medical world.

Figure 4. Diagram represents the steps in the design science research methodology.

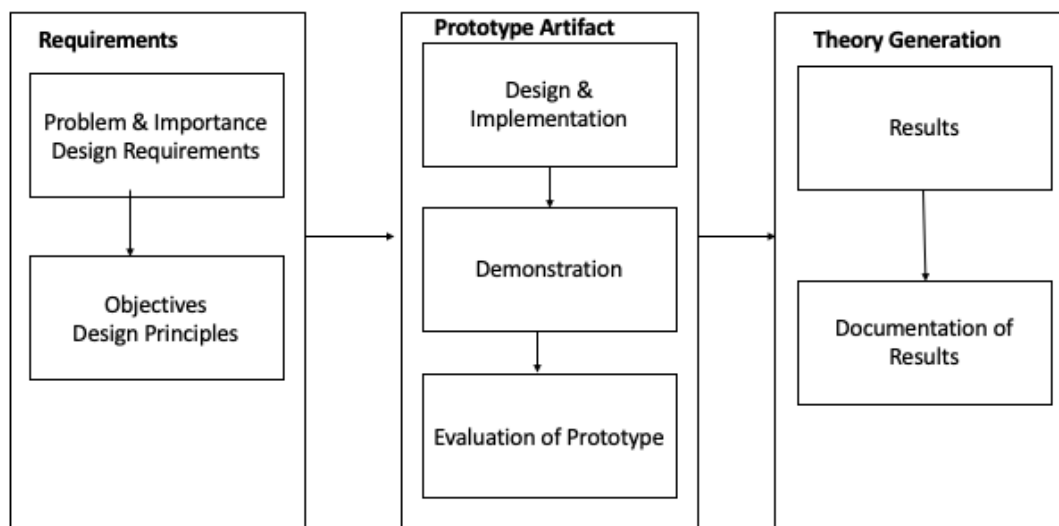


Table 1. Problem definition and objectives.

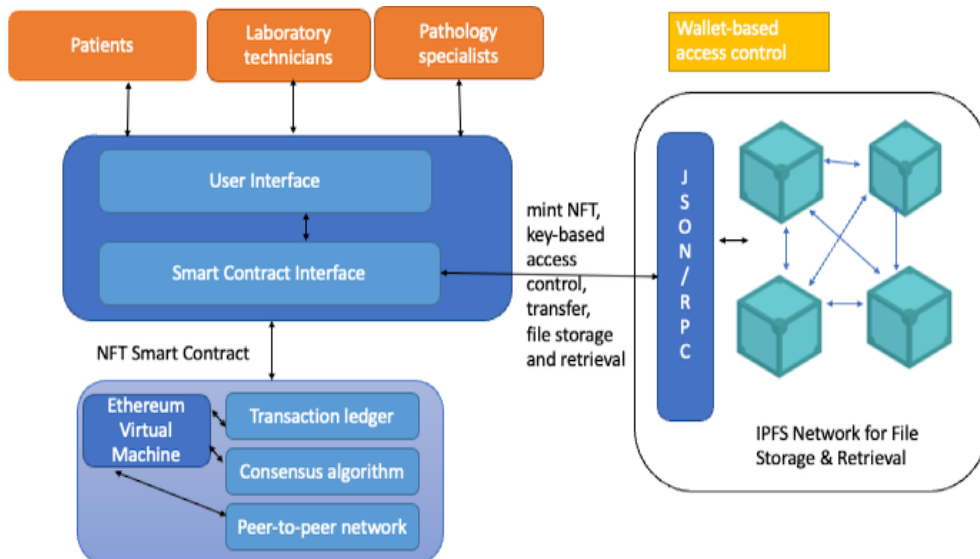
Problem definition	Importance and business outcome	Objectives of the solution
Long-term distributed and cost-effective storage of high-resolution images, metadata, and diagnoses provided by multiple users	Reduce the cost of adoption and utilization	Evaluation of alternate technologies that provide such storage; possibly separate the storage from the data transmission mechanism
Network transmission or the ability to share such high-resolution images and associated data	Improve speed of diagnosis for the disease	Mechanism to share data with others through the internet in a secure way
Data security	Improve security for all clinical data and provide access control	Mechanism to prevent unauthorized access of data
Data privacy	Improve the privacy of data in the system such that only the physician, pathologist, and patient can access these data	Mechanism to share the data on a need-to-know basis
Ability to maintain audit data transmission or data from shared access logs	Improve compliance using Health Insurance Portability and Accountability Act, General Data Protection Regulation, and privacy laws	A mechanism to trace how data have been shared among different users; store of logs should be perennial
Ability to maintain accuracy for creating a high-quality training data set for future application to artificial intelligence or deep learning models	Better training data can be stored on the network, and such data can facilitate better machine learning and artificial intelligence prediction	The mechanism above should have the ability to be audited, verified, and analyzed by third parties independent of the system
Reduce cost to the patient	Public infrastructure with well-established cryptographic protocols can enable the solution to scale significantly; network transmission is avoided by design	Network-attached storage costs are reduced for the solution; telemedicine and remote diagnosis are enabled easily
Improve the quality of care for the patient	Once higher accuracy for training data is established, the quality of care for the patient would have improved significantly	Machine-assisted diagnosis is possible
Ability to separate artifact ownership from token access	The pathological data record should explicitly state that the patient is the data owner, regardless of other individuals accessing the record	The metadata always explicitly records the ownership of the digital image by storing the patient’s pseudonymous information such as ID or wallet address that are unique to the patient
Ability to remove access after a prespecified time interval as per the legal and regulatory conditions	After a prespecified interval, the record should automatically be burned for all nonowners; those health care professionals who received the record should no longer have access to it	This functionality provides the ability to remove access to specific records and adhere to health care data storage and retrieval laws

Architecture and Design Components

Figure 5 depicts the software architecture of the decentralized digital pathology system. The system architecture conforms to

the model-view-controller architecture pattern, and the system’s main components are described in the next section.

Figure 5. Software architecture of the digital pathology system with blockchain and IPFS used for tokenizing, storing, and sharing access to the digital outputs of the health care system. IPFS: Interplanetary File System; NFT: nonfungible token.



Description

We use a model-view-controller software architecture pattern [24] to prototype the digital pathology system capable of addressing the requirements described in the *Introduction* section. The model stores the data about image scans and uses the interfaces provided by the view and controllers. Data are stored on the IPFS [25] and are usually of two types:

1. High-resolution image files about scans and any modifications made by the pathologist or other professionals, such as notes and image markings, to identify specific disease patterns, and
2. The JSON files containing metadata about the scans stored on the web or the diagnosis provided by the physician.

The IPFS provides a decentralized store for data and is based on caching data where file access frequency provides users with performant file storage.

We use Solidity programming language for programming the smart contract, which runs on the Ethereum test network. We use several JavaScript toolkits, namely, web3.js, node.js, and essential libraries associated with HTML5 and bootstrap.css, to code the front end. For testing the prototype, we used the Ganache command line toolkit (CLI), the VSCode library extensions for Solidity and HTML5 and JavaScript, and the integrated development environment for programming using Remix smart contract editor. We use the Ethereum Kovan and Ropsten test networks to implement the prototype and to validate our solution against the objectives set out in this paper.

The user interface provides a visual interface for physicians, laboratory technicians, and patients to view the data that they own or share with their wallets. While physicians, laboratory technicians, and patients can access the same image addressed using the CID on the IPFS, the main controlling file is the JSON, which interacts with the smart contract infrastructure and mints the token. The address stored (ie, the CID of the JSON file that contains metadata about the image or diagnosis) in the minted token is hidden from public access by encrypting the CID with the user's keys. The security aspects of the proposed system are described in the following section. The user interface then interacts with a controller consisting of a series of smart contract methods that mint new tokens, transfer tokens from the laboratory technician to the physician and from one physician to another using the NFT protocol, and so on. The decentralized application is responsible for creating newer images and minting the NFTs, which can potentially generate notes from physicians, clinical pathologists, or patients to enable data to be stored within the IPFS. This data model is described in the JSON file and indicates the different types of files used by such a system.

Because the IPFS stores each image file separately (as a unique addressable and unalterable hash), a JSON schema file similar to code listing 1 ([Multimedia Appendix 2](#)) creates a JSON file that contains the address for the CID for each NFT. Each time a new file is added, a new JSON metadata file is generated and stored within the IPFS. As a caveat, although the IPFS is a public infrastructure consisting of thousands of nodes from voluntary contributors, token-based decentralized storage mechanisms such as FileCoin and Storj are also available,

providing decentralized storage and enabling application developers to access guaranteed distributed storage on the network. Both Storj and FileCoin incentivize users to participate in decentralized file storage by rewarding contributors with cryptocurrency. The price for Storj is approximately US \$4 per month for 1-TB storage and US \$7 per month per TB of bandwidth, which is lower than most other web hosting or cloud service provider costs, thereby reducing storage and network transmission costs for high-resolution images.

From code listing 1, that is, the JSON metadata file, we obtain a unique CID that addresses each image location along with the pseudonymity of the owner of the image in the form of the owner's name and wallet ID. This information is unique and used to separate ownership from access rights. The blockchain provides the encryption and security for this JSON metadata file of the images used in the data set.

The CID of the metadata file is used as the input to the NFT mint (ERC-721) and provides the originator of the file access to this token. This token can then be transferred to other individuals on the network by simply using their wallet addresses. Every JSON metadata file associated with the record should contain the owner name and ID, and the patient's public wallet address that uniquely identifies the patient. When the metadata file must be shared among multiple people, the sender will have to mint multiple tokens with separate metadata files (CIDs) to transfer the token to other individuals. However, the base file will remain the same, and the image data file is referenced in each JSON file independently. The source code for minting the NFT in Solidity is listed in code listing 2 ([Multimedia Appendix 2](#)).

The original NFT can be transferred by sending the token to the recipient's wallet address using the *SafeTransfer* function described in code listing 3 ([Multimedia Appendix 2](#)).

Decentralized Application Functions to Mint, Store, Transfer Ownership of, and Expire (Burn) the Token

In the decentralized app, we use the smart contract code of the NFT standard (ERC-721) to mint an NFT on the Ethereum network and use the JSON file's IPFS URL (code listing 4 in [Multimedia Appendix 2](#)). The web interface invokes the smart contract from the Ethereum remote and internet-based machine to transfer the token from the source address to the destination address. Further, we transfer the token ownership to a receiver's blockchain wallet. This mechanism ensures that only the NFT owner has access to the JSON metadata file, which points to the IPFS URI of the pathological image sample. Further, the web interface allows us to create new files for diagnosis, upload new image files, mint new NFTs, and transfer ownership of NFTs, all of which originate and require the user to access a cryptocurrency wallet such as MetaMask. Once the user signs into the wallet, the wallet provides multilayer security for accessing original content stored within the IPFS and access the application. Within each wallet is access to the NFT for which the current wallet has ownership. The NFT will contain a unique CID for the JSON metadata file, containing further pointers to scanned images, diagnostics from physicians, and prescriptions. Code listing 3 shows how the NFT is minted based on the user input.

Code listing 5 (Multimedia Appendix 2) shows how to access the smart contract on the remote and internet-based machine from the front-end JavaScript of the decentralized application. We use Ethereum events and event listeners within the web3.js implementation standard to invoke the mint functionality of the smart contract. Code listing 5 shows that MintURI is the JavaScript function that captures and returns the token ID once the token is minted. MintURI retrieves the token ID and associates the token ID to the IPFS CID of the JSON file, which contains the metadata (as described in code listing 1).

Regarding privacy of data, this CID can be encrypted by a key pair stored in the wallet by the application when minting to give only the owner access to the URI where the JSON file is stored.

Transferring Ownership of the Pathological Record

In the code snippet (code listing 6 in Multimedia Appendix 2), we see the URI, the JSON file’s CID being transferred to the MedContract.methods.mint(...). The token is associated with the wallet (Ethereum) address 0x138a93...A8. Overall, once the token is minted, the owner can securely transfer the NFT over the network to another user. This transfers both the owner of the diagnosis file and the URI to the receiver. However, the actual data stored on the IPFS has never moved. The only transfer happening is the token’s ownership (which points to the CID of the JSON file data), which was initially associated with the sender’s wallet address. It will now be accessible only to the receiver. Code listing 6 describes the main functionality and how the user interface sends the code to other users.

The code listing 6 shows how data ownership is transferred from the senders to the receiver. The underlying CID (IPFS) and its contents can now be accessed by the receiver, who then examines the high-resolution image and either contributes by creating a new diagnosis file or creates a new image for minting the NFT.

Creating an Expiration Date–Driven NFT Access Control

One of the key requirements is for the token to expire as per the expiry date. Different state and local laws dictate how long each hospital can store a patient record and conditions around the storage of the record. For example, in the health care system that the coauthor of this paper works for, the hospital system stores the records (glass slides containing high-resolution microscopic scans and digitized images of these scans) for 3 years, after which they are securely discarded from their storage. However, the owner or patient can choose to retain their copy for as long as they choose.

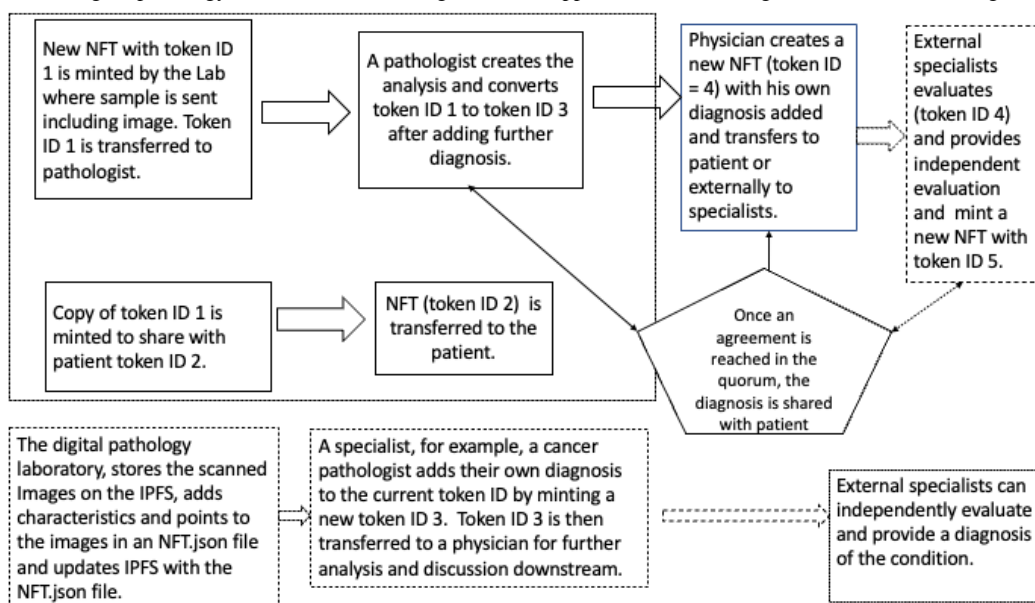
The approach we chose to replicate this functionality on the blockchain is as follows: The smart contract code supports the burn functionality for the token, which subsequently makes it inaccessible by any of the system’s users. Then, the web3.js application scans each record for the creation date and the owner’s wallet ID and burns those for all those token owners whose wallet IDs do not match the owner’s wallet ID, which is stated in the JSON record. Code listing 7 (Multimedia Appendix 2) displays the burn functionality for the token ID when the token was burned.

Upon execution of the contract within each user’s (eg, physician’s or even pathologist’s) decentralized application, those token ID’s with creation dates exceeding the expiration date would be burned and inaccessible to anyone else on the system. Only the record owner (or patient) would never execute the burn function, preserving original access to the NFT.

Use Cases Supported by Design and Implementation of the System

Several use cases are supported by the design and implementation of our system. One such user flow is shown in Figure 6, which directly refers to the aforementioned design requirements.

Figure 6. Illustration of a digital pathology workflow with nonfungible tokens applied to scanned images overall. NFT: nonfungible token.



Securing Data Storage for Privacy and Data Control

Access control, ownership, and confidentiality of data are three challenges in storing health care–related data on blockchain platforms. Because data reside on the blockchain and either the data or pointer to the data is publicly accessible and available to all users, how can users ensure that their data are accessed only by authorized individuals and that others do not have access to it? Furthermore, how do users ensure an audit trail for the accessed data, as auditability is one of the legal requirements from Health Insurance Portability and Accountability Act or General Data Protection Regulation? Similarly, how are ownership rights of data separated from the right to share the data with health care professionals for specific contexts of diagnosis?

To solve these challenges, we describe and implement the prototype for the security architecture in [Table 2](#) in the context of our digital pathology system on a public blockchain. The security architecture of the digital pathology system is multilayered based on the different access requirements in the system. The principal fundamental functions of the security system include authentication, data integrity, confidentiality, notarization or signature, access control, assurance of availability, and ownership. By authentication, we mean that the data are accessed only by those individuals who handle the data. Data integrity means that the data are never tampered with

by anyone in the system. The confidentiality property alludes to the fact that the actual data are not exposed to others, although pointers to the data reside in the blockchain. By notarization, we ensure that every time the data are transmitted or accessed and restored on the system, there is a valid signature of the individual who has accessed the record stored on the system. By access control, we mean that data access is tracked in the system, and only those allowed to access the data through a valid identifier can access the data. Finally, by assurance and availability, we ensure that data are available only to those individuals who have the right to know legally, and the security mechanism is available to all participants in the network (ie, the rules for access control and security are available to all users uniformly). The unique aspect of data record ownership is now added to these properties to the blockchain. The data record ownership ensures that only the primary owner of the data record, that is, the patient, has perennial access to the NFT. In addition, for all those who are not primary owners of the health care record (ie, the nonpatients) to whom the NFT has been transferred, the application would automatically invoke the smart contract's burn functionality after a preassigned period.

In [Table 2](#), we discuss the different levels of data security and privacy enforcement at three layers, namely, the application layer, smart contract layer, and data storage layer, that is, the IPFS layer for such a system.

Table 2. Security layers, property of the security architecture, and implementation.

Layers of the stack	Property of the security architecture	Implementation
Application layer and smart contract layer	Authentication	Users who require access to data are authenticated by their wallet (which contains a hash of the user's public key and a private key). Each time a message is sent, it is signed and verified using ECDSA ^a . Only those users who can sign in to their wallets can access their corresponding data in the form of the NFT ^b .
IPFS ^c layer and smart contract layer	Data integrity	Data integrity is managed and maintained by the underlying blockchain and IPFS layers. One of the properties of the blockchain is that the data cannot be manipulated ever. Similarly, the original image or diagnosis file stored in the IPFS cannot be altered. Changing the file will give rise to a new CID ^d , which will need minting a new token with different access controls.
Application layer for encryption and smart contract layer for storing the encrypted CID on the blockchain	Confidentiality of data	Although data stored on the blockchain is public information, the NFT being minted is minted off a JSON file's CID stored in the IPFS. The application can encrypt and store the encrypted CID on the block in the associated NFT accessible only to the wallet owner. The encryption is done by the application and not by the blockchain or the blockchain's smart contract. Each time a new token is minted, the resulting metadata file is uploaded onto the IPFS. The CID of the metadata file is encrypted by the application and stored on the blockchain as part of the contract.
Smart contract layer	Notarization	After signing with the owner's public key, each piece of information is notarized and stored on the blockchain. Each time the NFT is transferred to a new owner, at the application layer, the NFT is signed by the new owner's public address and later encrypted using the new owner's keys for storage on the blockchain at the application level, so that only the new owner can access the contents of the NFT.
Application layer and smart contract layer	Access control	Access control for a token is currently maintained by means of a wallet (both public and private keys). However, this access control can be maintained and moderated by the user.
Blockchain and IPFS layer	Availability	This is moderated by the underlying blockchain and IPFS infrastructure that has a 99.99% availability. The only limitation is network availability, which controls the rate at which data can be deposited and pulled from the network.

^aECDSA: Elliptic Curve Digital Signature Algorithm.

^bNFT: nonfungible token.

^cIPFS: Interplanetary File System.

^dCID: content identifier.

Results

After proof-of-concept implementation, we test and describe how the main objectives of digital pathology solutions are met. Furthermore, using our solution, we evaluate how this solution addresses the challenges associated with storage, transmission, cost-effectiveness, and security concerning digital pathology. [Figure 7](#) shows the screenshots of the prototype digital pathology and the system's user interface. In addition, these screenshots demonstrate how such a system works and the mode of operation of the digital pathology system.

In [Figure 7](#), we see how these images are accessible to the owner of the data set and now surface at the user interface. The pathologist and other experts who have access to the image

through the assignment of the NFT can now click on each image, open the image, and analyze the image through visual inspection or by the assistance of various artificial intelligence algorithms. The respective professional can then save the modified image and diagnosis onto the IPFS, mint a new NFT, and share it with the patient or other specialists in the group. Similarly, access to all such images occurs by signing in to the corresponding wallet of the user. Once signed in, the user has access to the corresponding NFT through a wallet. Such a design that separates the user actions from storage and records each action atop the blockchain creates a valuable chain of record. [Table 3](#) presents the benefits of our design vis-à-vis the original requirements, the implementation details, and the benefits of such a solution.

Figure 7. User interface of the digital pathology system that displays 4 specific scans the corresponding physician has access to.

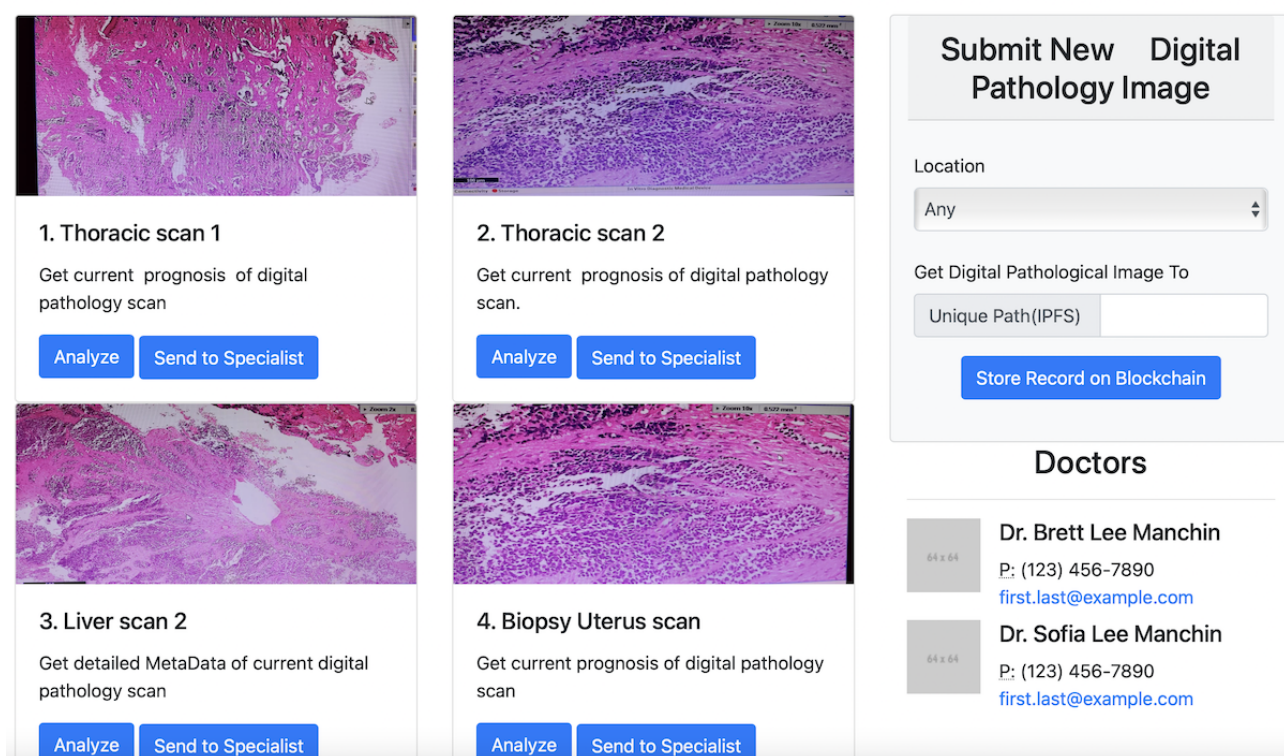


Table 3. Solution requirements and how they are addressed in the implementation.

Requirement	Implementation details	Benefits
Security	There are 2 layers of security for data in such a system. First, the IPFS ^a that stores images provide a high-security level by encrypting and splitting data into chunks stored on the network. Data are accessible only via a CID ^b , which is a large 256-byte hex code. Next, the JSON metadata file, which addresses the CID that is encrypted within the blockchain, is only accessible by those who use their keys to access the contract address where the data are stored. Refer to Table 2 for the implementation details.	The 2-layer security of smart contract data will only enable the wallet owner to access the record's contents. IPFS's CID is impossible to guess randomly.
Privacy of data	Only the user who has access to the corresponding wallet can access the data on the blockchain, and all other users will be blind to the use of data. Refer to Table 2 for an explanation of how data confidentiality and privacy are handled in the design.	This feature of the blockchain provides users (patients), physicians, and others the additional layer of security.
Low-cost, high-fidelity file storage	The IPFS is a public infrastructure secured by nodes running globally. The smart contract code can communicate with the IPFS if contracts prefer pinning (or converting to static storage) the data. An alternate public storage mechanism is FileCoin, which builds an incentive mechanism atop the IPFS to reimburse users for providing high-end storage and availability.	There is no need for compression of data or the manipulation of original data files for long-term storage. Such a storage method enables storage of the original files at high fidelity for further analysis.
High performance	Data and smart contract access is enabled to the high-throughput data storage via the IPFS and metadata file that contains a data pointer (ie, CID of the actual image). We also refer to other decentralized storage systems such as FileCoin and Storj, which are paid alternatives to the IPFS wherein crypto tokens enable the QoS ^c layer to ensure data availability. As a result, the actual file is never moved on the network, and only access controls to the CID containing JSON files are altered.	Data transfer does not occur on the network, except for the image being scanned and stored on the IPFS. Data access is provided via metadata files that are minted into NFTs. The NFTs, once transferred on the network, transfer the ownership of the underlying token to the receiver. As a result, the only piece of data transferred on the blockchain is the token through a transaction. This improves the performance of any system compared with a networked system infrastructure or a permissioned database structure where entire data files will have to move on the network.
Low transmission cost	Data transmission is accomplished by just transferring the token ownership on the blockchain corresponding to the file. Therefore, the original file is not transmitted over the network; instead, the JSON file's CID ownership is transferred to a different wallet on the blockchain.	The network transmission of the IPFS record means that only the minted token is transferred in ownership. The details of the token ID are stored and encrypted on the blockchain. Therefore, no actual data are transferring on the blockchain, reducing and improving bandwidth significantly.
Improve data accuracy	Every time the data are modified by the physician or other intermediaries, newer version-controlled files are created on the IPFS.	Higher accuracy of the data helps train artificial intelligence and machine learning models to improve prediction accuracy.
Ownership	Every specific period when NFT ^d ownership expires for legal reasons, such access is removed automatically through the smart contract's burn functionality. Only patients who own the token obtain a chance to hold onto the token and access-associated digital images. The rest of the owners will not have access after the expiry (burn of the token).	The ownership of the digital asset is separated from the functionality of the smart contract. Although the NFT can be shared with multiple practitioners, the actual access is determined by the legal policy of the health care system for nonpatient access. Only patients can access their own NFTs after the record's expiry date, where the NFT access is restricted for all others.

^aIPFS: Interplanetary File System.

^bCID: content identifier.

^cQoS: quality of service.

^dNFT: nonfungible token.

Discussion

Principal Findings

In this study, we introduced several innovations in decentralized digital pathology that help offset the costs of imaging, sharing, and distributing high-resolution slide scans for medical

diagnosis. Using private infrastructure mechanisms responsible for the maintenance of nodes and consensus mechanisms and administration of such systems controlled by a few firms would defeat the entire purpose of using digital pathology to improve the speed of diagnosis and the cost of patient care. Although other health care applications have discussed frameworks to store health care data on the blockchain or to use permissioned

networks with different types of encryptions, such a design would increase health care costs and system expenses. In addition, with private blockchains or consortium blockchains, the cost for maintaining the infrastructure and governance would centralize the blockchain system.

Although, on private permissioned blockchain, it is feasible to implement such on-chain storage without additional costs for storage, our design uses decentralized public storage as the back end, combined with established NFT standards. At the same time, private or permissioned blockchain architectures could lead to incompatibility among different chains and prevent external sharing of data because private blockchain with on-chain storage operates within network boundaries. Essentially, there would be a need to implement such permission blockchains across the hospital supply chains, including laboratories, partners, verification agencies, and clinics. These costs would prohibit the mass adoption of digital pathology, whose promise is to reduce costs, increase speeds, and use machine-assisted artificial intelligence and machine learning algorithms to communicate with others. In addition, software and hardware upgrades, technology obfuscation, network speeds are left to the blockchain maintainers in the case of private blockchains. As a result, a trade-off between private permission and consortium blockchains without permission would make a design such as the one presented here suitable. The public maintains the infrastructure with a consortium-driven networks and private and public networks.

We relied on a public blockchain infrastructure combined with decentralized storage in our solution, which we believe will drastically reduce the costs for storage transmission and provide a privacy-enabled and secure system for public health care. At the same time, the participants can also contribute to such infrastructure by offering nodes on which they could pin their

files for the IPFS. Alternatively, an inexpensive decentralized storage system such as Storj could be used, providing permissioned access to ample storage at a very affordable cost. The only piece of data that the blockchain stores is the encrypted CID for the IPFS metadata file. Although the IPFS metadata file is addressable and accessed outside the blockchain by users whose wallets have the authority to access the contents (image scans, diagnosis, treatment pathways, and other associated metadata), our security architecture prevents the public or other users from accessing the contents, thus providing security.

We reduced the cost of storage of high-resolution images by proposing the use of the IPFS, a secure and distributed peer-to-peer network with high availability. We used cryptocurrency wallets and smart contracts to create NFTs using the widely accepted ERC-721 standards. This design precludes the need to add a separate security layer for data storage, transmission, and access. Although we separated the storage of the actual image from its metadata, the transmission of the image entails changing the permission for accessing the metadata file through encryption with the receiver's keys. Our design approach does not require moving the entire high-resolution image to a different address on a network, which would likely consume much more network bandwidth and is inefficient, although this can also be accomplished.

Conclusions

Overall, the data transmission, data storage, and use of the ERC-721(NFT) standard for pathology-based image storage will improve the prospects for mass adoption of digital pathology. Such an architecture can significantly reduce the cost of the infrastructure required for digital pathology and, over time, improve the speed for diagnosis and affordable medical care.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questions answered by pathologist physician coauthor (SS) that discusses challenges with digital pathology.

[[DOCX File , 14 KB - jmir_v24i3e34207_app1.docx](#)]

Multimedia Appendix 2

Source code listings.

[[DOCX File , 18 KB - jmir_v24i3e34207_app2.docx](#)]

References

1. Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: challenges and opportunities. *Med Image Anal* 2016 Oct;33:170-175 [[FREE Full text](#)] [doi: [10.1016/j.media.2016.06.037](https://doi.org/10.1016/j.media.2016.06.037)] [Medline: [27423409](https://pubmed.ncbi.nlm.nih.gov/27423409/)]
2. Pantanowitz L. Digital images and the future of digital pathology. *J Pathol Inform* 2010 Aug 10;1(1):15 [[FREE Full text](#)] [doi: [10.4103/2153-3539.68332](https://doi.org/10.4103/2153-3539.68332)] [Medline: [20922032](https://pubmed.ncbi.nlm.nih.gov/20922032/)]
3. Schmitt M, Maron RC, Hekler A, Stenzinger A, Hauschild A, Weichenthal M, et al. Hidden variables in deep learning digital pathology and their potential to cause batch effects: prediction model study. *J Med Internet Res* 2021 Feb 02;23(2):e23436 [[FREE Full text](#)] [doi: [10.2196/23436](https://doi.org/10.2196/23436)] [Medline: [33528370](https://pubmed.ncbi.nlm.nih.gov/33528370/)]
4. Tizhoosh H, Pantanowitz L. Artificial intelligence and digital pathology: challenges and opportunities. *J Pathol Inform* 2018;9(1):38 [[FREE Full text](#)] [doi: [10.4103/jpi.jpi_53_18](https://doi.org/10.4103/jpi.jpi_53_18)] [Medline: [30607305](https://pubmed.ncbi.nlm.nih.gov/30607305/)]

5. Jahn SW, Plass M, Moinfar F. Digital pathology: advantages, limitations and emerging perspectives. *J Clin Med* 2020 Nov 18;9(11):3697 [FREE Full text] [doi: [10.3390/jcm9113697](https://doi.org/10.3390/jcm9113697)] [Medline: [33217963](https://pubmed.ncbi.nlm.nih.gov/33217963/)]
6. Hartman D, Pantanowitz L, McHugh J, Piccoli A, OLeary M, Lauro G. Enterprise implementation of digital pathology: feasibility, challenges, and opportunities. *J Digit Imaging* 2017 Oct 23;30(5):555-560 [FREE Full text] [doi: [10.1007/s10278-017-9946-9](https://doi.org/10.1007/s10278-017-9946-9)] [Medline: [28116576](https://pubmed.ncbi.nlm.nih.gov/28116576/)]
7. Romero Lauro G, Cable W, Lesniak A, Tseytlin E, McHugh J, Parwani A, et al. Digital pathology consultations-a new era in digital imaging, challenges and practical applications. *J Digit Imaging* 2013 Aug 29;26(4):668-677 [FREE Full text] [doi: [10.1007/s10278-013-9572-0](https://doi.org/10.1007/s10278-013-9572-0)] [Medline: [23359091](https://pubmed.ncbi.nlm.nih.gov/23359091/)]
8. Zuo Z, Lan X, Deng L, Yao S, Wang X. An improved medical image compression technique with lossless region of interest. *Optik* 2015 Nov;126(21):2825-2831. [doi: [10.1016/j.ijleo.2015.07.005](https://doi.org/10.1016/j.ijleo.2015.07.005)]
9. Kempfert A, Reed B. Health care reform in the United States: HITECH Act and HIPAA privacy, security, and enforcement issues. *FDCC Quarterly* 2011;61(3):240-273 [FREE Full text]
10. Nahra KJ. HIPAA security enforcement is here. *IEEE Secur Privacy Mag* 2008 Nov;6(6):70-72. [doi: [10.1109/msp.2008.143](https://doi.org/10.1109/msp.2008.143)]
11. Janowczyk A, Madabhushi A. Deep learning for digital pathology image analysis: a comprehensive tutorial with selected use cases. *J Pathol Inform* 2016;7(1):29 [FREE Full text] [doi: [10.4103/2153-3539.186902](https://doi.org/10.4103/2153-3539.186902)] [Medline: [27563488](https://pubmed.ncbi.nlm.nih.gov/27563488/)]
12. Domingue J, Third A, Ramachandran M. The FAIR TRADE framework for assessing decentralised data solutions. In: *Proceedings of the Companion Proceedings of the 2019 World Wide Web Conference*. 2019 Presented at: Companion Proceedings of the 2019 World Wide Web Conference; May, 2019; San Francisco, USA. [doi: [10.1145/3308560.3317076](https://doi.org/10.1145/3308560.3317076)]
13. Subramanian H. Decentralized blockchain-based electronic marketplaces. *Commun ACM* 2017 Dec 27;61(1):78-84. [doi: [10.1145/3158333](https://doi.org/10.1145/3158333)]
14. Costello SS, Johnston DJ, Dervan PA, O'Shea DG. Development and evaluation of the virtual pathology slide: a new tool in telepathology. *J Med Internet Res* 2003 Jun 13;5(2):e11 [FREE Full text] [doi: [10.2196/jmir.5.2.e11](https://doi.org/10.2196/jmir.5.2.e11)] [Medline: [12857667](https://pubmed.ncbi.nlm.nih.gov/12857667/)]
15. Park YR, Lee E, Na W, Park S, Lee Y, Lee J. Is blockchain technology suitable for managing personal health records? Mixed-methods study to test feasibility. *J Med Internet Res* 2019 Feb 08;21(2):e12533 [FREE Full text] [doi: [10.2196/12533](https://doi.org/10.2196/12533)] [Medline: [30735142](https://pubmed.ncbi.nlm.nih.gov/30735142/)]
16. Maslove DM, Klein J, Brohman K, Martin P. Using blockchain technology to manage clinical trials data: a proof-of-concept study. *JMIR Med Inform* 2018 Dec 21;6(4):e11949 [FREE Full text] [doi: [10.2196/11949](https://doi.org/10.2196/11949)] [Medline: [30578196](https://pubmed.ncbi.nlm.nih.gov/30578196/)]
17. Ichikawa D, Kashiyama M, Ueno T. Tamper-resistant mobile health using blockchain technology. *JMIR Mhealth Uhealth* 2017 Jul 26;5(7):e111 [FREE Full text] [doi: [10.2196/mhealth.7938](https://doi.org/10.2196/mhealth.7938)] [Medline: [28747296](https://pubmed.ncbi.nlm.nih.gov/28747296/)]
18. Khurshid A. Applying blockchain technology to address the crisis of trust during the COVID-19 pandemic. *JMIR Med Inform* 2020 Sep 22;8(9):e20477 [FREE Full text] [doi: [10.2196/20477](https://doi.org/10.2196/20477)] [Medline: [32903197](https://pubmed.ncbi.nlm.nih.gov/32903197/)]
19. Sylim P, Liu F, Marcelo A, Fontelo P. Blockchain technology for detecting falsified and substandard drugs in distribution: pharmaceutical supply chain intervention. *JMIR Res Protoc* 2018 Sep 13;7(9):e10163 [FREE Full text] [doi: [10.2196/10163](https://doi.org/10.2196/10163)] [Medline: [30213780](https://pubmed.ncbi.nlm.nih.gov/30213780/)]
20. Subramanian H, Liu R. Blockchain and smart contracts. *J Database Manag* 2019 [FREE Full text]
21. What is ERC-721? ERC-721. URL: <http://erc721.org/> [accessed 2022-03-24]
22. Sengupta A, Subramanian H. User control of personal mHealth data using a mobile blockchain app: design science perspective. *JMIR Mhealth Uhealth* 2022 Jan 20;10(1):e32104 [FREE Full text] [doi: [10.2196/32104](https://doi.org/10.2196/32104)] [Medline: [35049504](https://pubmed.ncbi.nlm.nih.gov/35049504/)]
23. Peffers K, Tuunanen T, Rothenberger MA, Chatterjee S. A design science research methodology for information systems research. *J Manag Inf Syst* 2014 Dec 08;24(3):45-77. [doi: [10.2753/mis0742-1222240302](https://doi.org/10.2753/mis0742-1222240302)]
24. Model-view-controller pattern. In: *Learn Objective-C for Java Developers*. New York City: Apress; 2009.
25. Kumar R, Tripathi R. Blockchain-based framework for data storage in peer-to-peer scheme using interplanetary file system. In: *Handbook of Research on Blockchain Technology*. Cambridge, Massachusetts, United States: Academic Press; Jan 1, 2020.

Abbreviations

- CID:** content identifier
- DSRP:** design science research process
- ERC:** Ethereum Request for Comments
- IPFS:** Interplanetary File System
- NFT:** nonfungible token
- RQ:** research question

Edited by A Mavragani; submitted 14.10.21; peer-reviewed by C Reis; comments to author 05.11.21; revised version received 31.01.22; accepted 01.03.22; published 28.03.22.

Please cite as:

Subramanian H, Subramanian S

Improving Diagnosis Through Digital Pathology: Proof-of-Concept Implementation Using Smart Contracts and Decentralized File Storage

J Med Internet Res 2022;24(3):e34207

URL: <https://www.jmir.org/2022/3/e34207>

doi: [10.2196/34207](https://doi.org/10.2196/34207)

PMID: [35343905](https://pubmed.ncbi.nlm.nih.gov/35343905/)

©Hemang Subramanian, Susmitha Subramanian. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 28.03.2022. 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.

Original Paper

A Web-Based Self-assessment Model for Evaluating Multidisciplinary Cancer Teams in Spain: Development and Validation Pilot Study

Mercedes Guilabert¹, PhD; Joan Prades^{2,3}, PhD; Josep M Borrás^{2,3}, MD, PhD; Inmaculada Maestu⁴, MD, PhD; Juan Antonio Guerra⁵, MD; Lluís Fumadó⁶, MD, PhD; José Joaquín Mira^{1,7,8}, PhD; AEMAC Program Research Team⁹

¹Health Psychology Department, Miguel Hernandez University, Elche, Spain

²Department of Health, Catalanian Cancer Strategy, Barcelona, Spain

³Department of Clinical Sciences, University of Barcelona, Instituto de Investigación Biomédica de Bellvitge, Barcelona, Spain

⁴Medical Oncology Service, Doctor Peset University Hospital, Valencia, Spain

⁵Department of Hematology and Oncology, Fuenlabrada University Hospital, Fuenlabrada, Spain

⁶Urological Cancer Functional Unit, Hospital del Mar-Parc de Salut Mar-IMIM, Barcelona, Spain

⁷Atenea Research Group, Foundation for the Promotion of Health and Biomedical Research, Sant Joan d'Alacant, Spain

⁸Alicante-Sant Joan Health Department, Alicante, Spain

⁹see Authors' Contributions

Corresponding Author:

Mercedes Guilabert, PhD
Health Psychology Department
Miguel Hernandez University
Avda de la Universidad s/n
Elche, 03202
Spain
Phone: 34 96 665 83 17
Email: mguilabert@umh.es

Abstract

Background: Tumor boards constitute the main consensus and clinical decision-making body of multidisciplinary teams (MDTs) in cancer care. With the increasing clinical complexity of treatment options (eg, targeted therapies, multimodal treatments) and the progressive incorporation of new areas of intervention (eg, survivorship care), tumor boards are now required to play a central role in all cancer processes. However, although frameworks are in place to evaluate MDT quality, only few web-based tools are available for this purpose; indeed, no web-based MDT evaluation tools have been developed for or adapted to the Spanish National Health System.

Objective: The first aim of this study was to develop a web-based self-assessment model (Autoevaluación de Equipos Multidisciplinares de Atención al Cáncer [AEMAC]) for evaluating multidisciplinary cancer teams in Spain and the second aim was to validate this tool by testing its metric properties, acceptability, and usability.

Methods: We designed and validated the AEMAC program in 3 stages. In the first stage (research), we reviewed the available scientific evidence and performed a qualitative case study of good practice in multidisciplinary care within the Spanish National Health System (n=4 centers and 28 health care professionals). The results were used to define the thematic areas and quality criteria for the self-evaluation model, which were then discussed and validated by a group of experts. The second stage (development) involved the technological development of a web app that would be accessible from any mobile device. In the third stage (piloting and validation), we conducted 4 pilot tests (n=15 tumor boards, 243 professionals) and used the results to analyze the acceptability and usefulness of the tool.

Results: We designed a self-assessment model based on 5 thematic areas encompassing a total of 25 quality components, which users rated on a 3-option development scale. The evaluation process, which was managed entirely from the web app, consisted of individual self-assessment, group prioritization, and creation of an improvement plan. Cronbach alpha (.86), McDonald's omega (0.88), and various fit indices (comparative fit index between 0.95 and 1 and goodness-of-fit index between 0.97 and 0.99

for all 5 aspects) confirmed internal consistency. The mean rating for overall satisfaction with the tool and for consistency between the content of the tool and the reality of tumor boards was 7.6 out of 10.

Conclusions: The results obtained during the period of research and piloting of the AEMAC program showed that it has an appropriate structure and metric properties and could therefore be implemented in a real context and generalized to other hospitals. As a virtual tool, it helps to measure the key aspects of MDT quality, such as effectiveness of collaboration and communication, leadership, and the organizational environment.

(*J Med Internet Res* 2022;24(3):e29063) doi:[10.2196/29063](https://doi.org/10.2196/29063)

KEYWORDS

web-based tool; multidisciplinary care; cancer; evaluation; quality assurance

Introduction

Background

Tumor boards constitute the main consensus and clinical decision-making body of multidisciplinary teams (MDTs) in cancer care [1]. The direct and indirect benefits of MDTs have been confirmed through extensive analysis [2-8], and the proper organization and systematic implementation of these bodies has become the central focus of cancer control policies [9]. With the increasing clinical complexity of treatment options (eg, targeted therapies, multimodal treatments) and the progressive incorporation of new areas of intervention such as attention to survival as well as health care objectives such as the reduction of waiting times, MDTs have come to assume a central role in cancer care [10].

Rationale of MDT Assessment

Tumor boards are generally based on regular work sessions that require the participation of a large number of professionals and departments, but hospitals often fail to define the internal organization of these boards, the scope of their work, or the resources needed. Tumor boards should be subject to analysis and intervention based on their importance for patients and the economic costs involved to assess their degree of usefulness and the quality of their decisions and to establish necessary changes. From this perspective, the multiple benefits they may provide do not inevitably result from a policy decision [11]. In the context of the Spanish National Health System (SNS), although the cancer strategy set out by the Ministry of Health defines MDTs as a reference framework for oncology services [12], the degree of MDT formalization and work capacity varies substantially between hospitals [13,14]. Moreover, although many scientific societies and European organizations have made attempts to strengthen MDTs by establishing quality criteria and work procedures, only few web-based tools are in place for evaluating the effectiveness of teamwork. The tools currently available, such as MDT-FIT (MDT-feedback for improving teamworking) [15], MDT-OARS (MDT-observational assessment scale) [16,17], MDT-QuIC (MDT-quality improvement rating scale) [18], and MDT-MODE (MDT-metric of decision making) [19] vary in terms of methodology (eg, checklist, observation) and availability as a virtual tool [20].

Routines and Gaps in the Work of Tumor Boards in Spain

The large number of professionals involved in cancer care and the complexity of the diagnostic therapeutic process mean that coordination problems and poor communication are very likely to occur. The evidence points to the importance of these aspects, which can contribute to improving the survival and quality of life of patients with cancer. Thus, the implementation of a developed model of multidisciplinary care is one of the greatest challenges to progress in improving the quality of care for patients with cancer. In Spain, within the context of the SNS, the complexity of multidisciplinary care and its impact on cancer care depends on many factors: the different organizational forms according to the territories, decision making in the diagnostic and therapeutic process, or the cultural resistance to change that working in a multidisciplinary manner may entail. Since there are many barriers to the implementation of multidisciplinary care, this interactive tool can enable the diagnosis of MDTs, thereby offering a map of interventions for the proper functioning of the teams. Other international experiences [15-19] highlight the fact of having an instrument to perform self-diagnosis; the tool presented here additionally makes it possible to have the tool available in a web-based format, which facilitates the rapid creation of a roadmap for the appropriate management of the tumor committee within the SNS.

Autoevaluación de Equipos Multidisciplinares de Atención al Cáncer Program

To promote multidisciplinary cancer care within the SNS, we developed a program for the self-assessment of MDTs in cancer care (AEMAC, by its acronym in Spanish [Autoevaluación de Equipos Multidisciplinares de Atención al Cáncer]). The AEMAC program is a web-based tool designed to facilitate the evaluation of critical components related to the internal structure, coordination and communication, organizational context, and development possibilities of MDTs. The objective of this study was to research, develop, and validate the AEMAC program as a quality self-assessment tool for MDTs.

Methods

This study was performed in 3 stages.

Research and Expert Assessment

The first stage, focused on the design of the tool, drew on 2 sources of information: the available scientific literature (including gray literature) and the experiences of tumor boards

from hospitals of varying structure, size, and capacity to attend patients with cancer.

First, we updated a systematic review that had been conducted in 2014 by 2 of the authors of this paper (JP and JMB), extending the search to November 2016 while maintaining the same search criteria [10]. We also consulted a set of strategic documents to analyze organizational principles and standards in multidisciplinary care [17,21-27]. We decided to base the first version of the AEMAC program on the document *Policy Statement on multidisciplinary cancer care* [21], which defines the core elements that all tumor-based MDTs should include. By synthesizing the information collected in this systematic review, we were able to draft the interview script to be used in the development stage.

Second, we performed a qualitative case study on practices of excellence in multidisciplinary care in Spanish public hospitals in the period spanning May 2017 to September 2017 [28,29]. We included tumor boards specializing in specific tumors of all prognoses and locations and more general boards (eg, digestive system tumors) with a structured MDT that had been operating for at least 3 years in SNS hospitals located in all regions of Spain. A further inclusion criterion was the presence of a clinical unit for centralizing the whole care process of specific pathologies. After the selection process, we planned on-site visits to study in detail the work of MDTs in the following hospitals: Hospital Universitario del Mar, Barcelona (lung cancer); Hospital Universitario Ramón y Cajal, Madrid (breast cancer); Complejo Hospitalario de Navarra (digestive system tumors); and Hospital Clínico de Salamanca (multiple cancers). This sample of hospitals was endorsed by the scientific

societies involved in the program (Table S1 of [Multimedia Appendix 1](#)). Additionally, we interviewed 20 professionals from the following specialties: medical oncology (n=4), radiation oncology (n=4), other medical-surgical specialties (n=4), pathological anatomy (n=2), radiology (n=4), and nursing (n=2). These visits and interviews were aimed at identifying the key thematic areas and components to include in the MDT self-assessment model.

The resulting preliminary matrix was assessed by 6 SNS experts who were selected according to the criteria of territorial representativeness, different specialties, levels of care, and scientific or management experience (as of September 2017). Specifically, they assessed the readability, face validity, and content of the tool according to the legibility of the identified components and coherence of the model and its elements; the adaptability of the components to the reality of the SNS; the weighting of the components according to their level of importance; and the definition of future components.

Information Technology Development and Phases in the Self-assessment Process

The second stage consisted of developing the web-based tool. The AEMAC program was conceived from the beginning as a web app that could be activated and managed from any mobile device. The web-based tool [30] was designed to facilitate consensus within the tumor board on areas for improvement and possible interventions and to establish the work schedule for achieving these improvements and ensuring their sustainability. The self-assessment was structured in 3 phases (see [Textbox 1](#), [Multimedia Appendices 2-4](#), and [Figure 1](#)).

Textbox 1. Phases in the Autoevaluación de Equipos Multidisciplinares de Atención al Cáncer program self-assessment process.

Phase 1: Individual self-assessment ([Multimedia Appendix 2](#))

- All board members independently fill out the web-based questionnaire.
- Each component has 3 evaluation possibilities, ranging from least development (first option) to greatest development (third option), with an intermediate category (second option).
- The evaluation process for each component is completed with a subquestion on the “possibility of improvement” if option 1 is chosen or on the “risk of worsening” if option 3 is selected. These items are evaluated based on a scale of 0 to 10 points, and this score is included in the prioritization algorithm. For example, components with poor ratings and without the possibility of improvement obtain a lower relative weight.

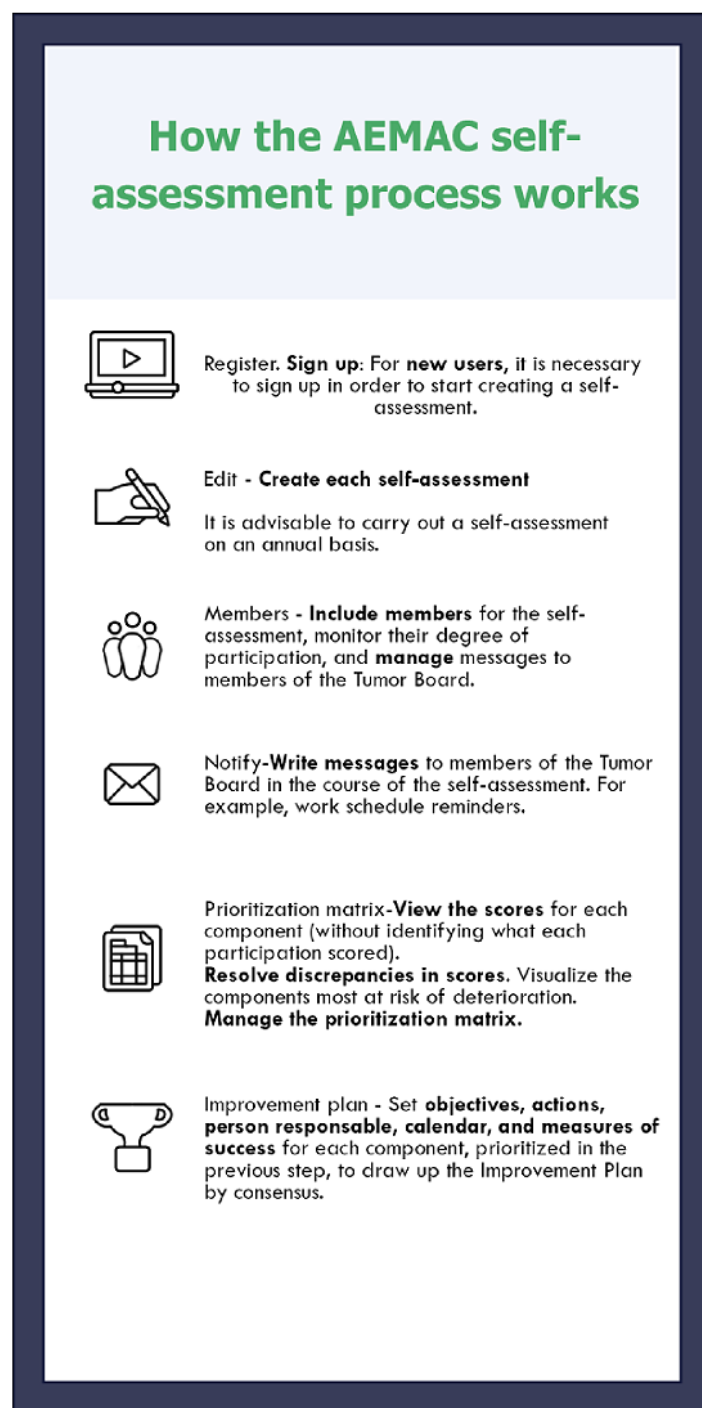
Phase 2: Group prioritization process ([Multimedia Appendix 3](#))

- The team (or a subgroup of 3 team members) assesses the results, resolves any team discrepancies that emerged in phase 1, and prioritizes areas for improvement.
- Discrepancies occur when none of the 3 response options accounts for at least 60% of the responses for any component.

Phase 3: Creating an improvement plan ([Multimedia Appendix 4](#))

- The team sets objectives and actions for an improvement plan spanning 1 to 2 years.
- Responsible parties are assigned and indicators are established so that the degree of achievement can be measured.

Figure 1. Working of the AEMAC self-assessment process. AEMAC: Autoevaluación de Equipos Multidisciplinares de Atención al Cáncer.



Data are stored and participants' identities are protected on the server at Miguel Hernández Public University to ensure anonymous data management. This is fundamental because the AEMAC program automatically shares compared information with other MDTs of the same hospital level and pathology to enhance benchmarking. The hospital level is captured through the postal code of the hospital that was provided during the initial registration of the self-assessment process.

Piloting and Validation

The third stage included a pilot study to test the adequacy of the procedure and to analyze the metric characteristics of the developed instrument. To test the acceptability and usefulness of the AEMAC program, pilot tests were performed in 4 SNS hospitals—Hospital Universitario del Mar (Barcelona), Hospital Universitario de Fuenlabrada (Madrid), Hospital Universitario Doctor Peset (Valencia), and Hospital Universitario La Paz (Madrid)—including a total of 15 tumor boards and 243

professionals. These hospitals belonged to 3 different regional health services and were representative of the different types of Spanish public hospitals. The number of beds ranged from 400 to 1000. Three researchers attended these meetings as observers, information technology consultants, and evaluators of the experience (JJM, MG, JP).

The first main objective of the pilot tests was to observe the clarity of the concepts used in each component and the consistency of the criteria used to define these concepts. These criteria include correlation (ie, if the tumor board hypothesized that good multidisciplinary work was being done in different areas, the self-assessment of the professionals on the tumor board should be in segment 3 of the areas they considered to be strengths), variability (ie, the second option should not be selected by default), and acceptability (ie, professionals should not feel penalized by “negative” language). The second aim of the pilot tests was to assess the usability and functionality of the website, considered as critical elements in the time for individual completion of the questionnaire and the general usability of the instrument (eg, interface user friendliness, ease of transition between phases). Additionally, the responses to each of the components were analyzed considering both floor and ceiling effects to rule out elements with little variability, elements that could not be implemented, and elements that were already fully implemented.

We also analyzed the degree of discrepancy in the evaluations of different members of the tumor boards for each component, acknowledging that systematic discrepancies between all participants of all boards must represent inappropriate wording. Finally, through confirmatory factor analysis, the coherence of the grouping of components by thematic area was analyzed using the comparative fit index, Jöreskog and Sörbom’s goodness-of-fit index, adjusted goodness-of-fit index, root mean square residual, standardized root mean square residual, and root mean square error of approximation. The model was accepted when comparative fit index, Jöreskog and Sörbom’s goodness-of-fit index, and adjusted goodness-of-fit index scores were greater than 0.95 and root mean square residual, standardized root mean square residual, and root mean square error of approximation scores were less than 0.08 [31,32]. The internal consistency of the resulting factorial solution was analyzed using Cronbach alpha and McDonald’s omega statistical indicators.

A total of 243 pilot study participants were asked to complete a survey to assess the critical issues of the tool ([Multimedia Appendix 5](#)). The anonymized responses were analyzed using descriptive statistics. The first part of the survey was designed to assess user experience of the AEMAC program and included 5 items with a response scale ranging from 0 (unsatisfactory) to 10 (totally satisfactory). The questions covered ease of answering the AEMAC questionnaire, coherence between questions, adequacy of the response options, and the method for establishing an improvement plan. The second part of the survey aimed to evaluate the adequacy of the established critical and semicritical components.

Ethics Approval

The study was approved by the Ethics and Research Integrity Committee of the Miguel Hernández University (AUT.DPS.JMS.02.21).

Results

Research and Expert Assessment

By the end of the research stage, we were able to define the following 5 thematic areas: (1) preparation and organization of the tumor board, (2) board decision-making process, (3) continuity of care, (4) organizational context, and (5) cross-disciplinary roles and team cohesion. These 5 thematic areas encompassed a total of 27 components considered to have a potential impact on quality of care. The following changes were made following the expert assessment of the preliminary matrix:

1. *Legibility of the identified components and coherence of the model and its elements.* Two components were added, and changes were made in the wordings to improve clarity and relevance and to avoid any overlaps.
2. *Adaptability of the components to the reality of the SNS.* It was decided to ensure adaptability of the AEMAC program to any MDT, regardless of hospital level (university, community, or local hospital).
3. *Weighting of the components according to their level of importance.* The experts agreed on different weights for the components according to the categories “normal,” “high,” or “very high” importance. This had an impact on the prioritization algorithm used during the self-assessment.
4. *Definition of future components.* Six components were defined with the prospect of being included in 5 years to replace outdated components.

The adjusted matrix can be seen in Table S2 of [Multimedia Appendix 1](#).

Information Technology Development and Phases in the Self-assessment Process

The AEMAC program was developed as a 3-phase process with the dual objective of gathering individual and aggregated perspectives on the quality of teamwork and encouraging dialogue, negotiation, and formulation of possibilities for improvement based on the critical aspects identified.

Piloting and Validation

The pilot study participants agreed that the website format and the fact that the questionnaire could be completed on any mobile device made the tool easy to use. Discrepancies could be resolved by a subgroup of professionals, which facilitated the process, although some teams had reservations in this regard. When faced with certain questions, some professionals stated that they required more information to give their opinion, usually because of lack of direct contact with patients. We, nonetheless, agreed to maintain the obligation of answering all the standards included in the self-assessment questionnaire. The time taken to complete the AEMAC questionnaire ranged from 12 to 14

minutes, which was considered satisfactory. The average group discussion time was 15 minutes for phase 2 (range, 10-20 minutes) and 30 minutes for phase 3 (range, 20-50 minutes). The aspect with the lowest acceptability rating was the registration process for initiating a self-assessment. MDTs in the SNS usually have no administrative support, which means registration has to be managed by the professionals themselves.

For most components of the thematic areas “Preparation and organization of the board” and “Continuation of the care process,” the third option (“Representing greater development”) was selected more frequently than either of the other two. Regarding “Board decision-making process” and “Organizational context,” the bulk of the responses fell between options 1 and 2. Finally, great variability in the responses was observed in “Cross-disciplinary roles and team cohesion”. Table S3 of [Multimedia Appendix 1](#) shows the combined scores from

all boards, indicating the range of scores observed and the response frequencies of each option for each component. The results of this first analysis identified no ceiling or floor effects to substantiate eliminating any of the 27 components. However, after reviewing the confirmatory factor analysis results, we eliminated the components “Patient information process” (from the thematic area “Board decision-making process”) and “Team-patient communication framework” (from “Cross-disciplinary roles and team cohesion”). We also moved the “Team cohesion” component to the thematic area “Board Preparation and Organization.” The fit indices of this model confirmed that the various components were well assigned to the thematic areas. [Table 1](#) shows the goodness-of-fit indices for each thematic area of the resulting model. The “Board decision-making process” obtained the best fit indices according to the previously defined benchmark.

Table 1. Confirmatory factor analysis adjustment indices.

Adjustment indices	Preparation/organization	Decision making process	Continuity of care	Organizational context	Cross-disciplinary roles
Comparative fit index ^a	0.95	1.00	0.96	0.96	0.99
Jöreskog and Sörbom’s adjusted goodness-of-fit index ^b	0.94	0.99	0.97	0.94	0.98
Jöreskog and Sörbom’s goodness-of-fit index ^b	0.97	0.99	0.99	0.98	0.99
Root mean square residual ^c	0.02	0.01	0.01	0.02	0.01
Standardized root mean square residual ^c	0.05	0.02	0.03	0.04	0.02
Root mean square error of approximation (90% CI)	0.05 (0-0.08)	0 (0-0.1)	0.04 (0-0.1)	0.08 (0.03-0.1)	0.03 (0-0.1)

^aRanges from 0 to 1, with a value of 0.9 being the minimum required to defend the model.

^bGoodness-of-fit index and adjusted goodness-of-fit index: range between 0 and 1 and those models that exceed 0.9 are considered adequate models.

^cRoot mean square error of approximation and standardized root mean square: indicators of a good fit with values less than 0.05.

[Table 2](#) shows the revised structure of the AEMAC program. The optimized model can be seen in [Figures 2-6](#) (optimized model of the questionnaire factorial structure based on confirmatory factor analysis from the validation study performed for each factor, factors 1-5, Structural Equation Modelling Path Diagram chart). The values shown in [Figures 2-6](#) represent the standardized solution of the confirmatory factor analysis equations. The values above the lines in the figures are the estimates of the regression coefficients of the common factors and of the specific factors, that is, the contribution of each item

to the factors, which is also called factor loadings. The R² is the variance explained by the factor for each of the items.

In the internal consistency analysis, the AEMAC program achieved a Cronbach alpha score of .86 and a McDonald’s omega score of 0.88. [Table 3](#) and [Table 4](#) show the combined responses of the survey evaluating the AEMAC program. In total, 40 out of 243 professionals responded (16.5% response rate). Seven (88%) of the 8 components that were considered critical and all 3 components that were considered semicritical received a high adequacy rating.

Table 2. Structure of the revised Autoevaluación de Equipos Multidisciplinares de Atención al Cáncer program following confirmatory factor analysis.

Thematic area	Items
Preparation and organization of the board	Attendance and representation
	Patient schedule
	Meeting frequency
	Cases discussed
	Involvement of professionals responsible for patients
	Time management efficiency
	Case presentation
	Team cohesion
Board decision-making process	Learning and updating knowledge
	Psychosocial perspective
	Oncogeriatric perspective
	Clinical trials
Continuity of care	Computerized record of decisions
	Decision implementation
	Follow-up planning
	Care for long-term survivors of cancer
Organizational context	Board time protection
	Administrative support
	Meeting room
	Technological resources
	Role of the Hospital Tumor Board
Cross-disciplinary roles	Board chair or coordinator
	Nursing case manager
	Key points in team-patient communication
	Team evaluation

Figure 2. Optimized factor 1 model based on the confirmatory factor analysis of the validation study carried out. The values above the lines are the estimates of the regression coefficients of the common factors and of the specific factors, that is, the contribution of each item to the factors, which is also called factor loadings. The R^2 is the variance explained by the factor for each of the items.

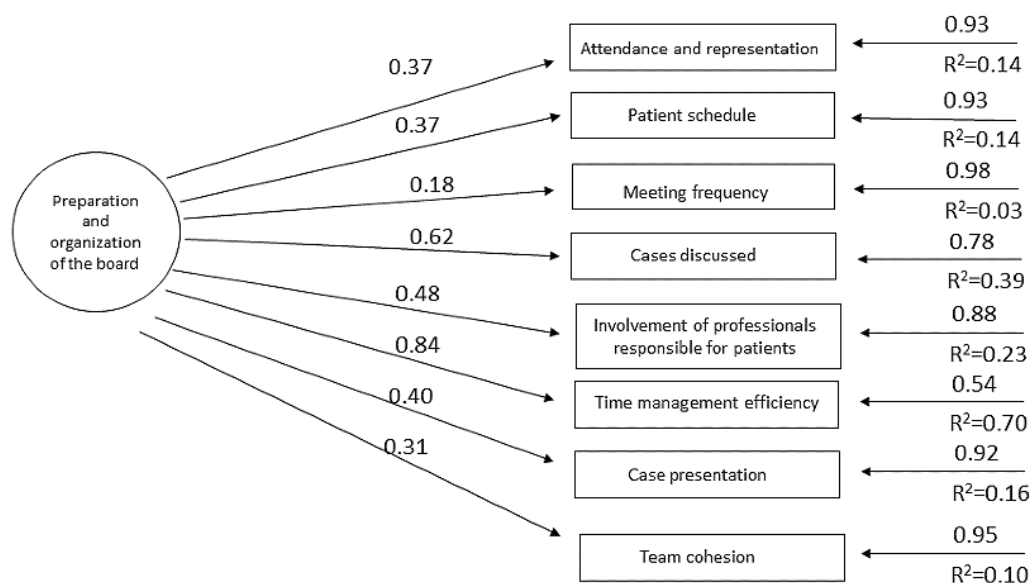


Figure 3. Optimized factor 2 model based on the confirmatory factor analysis of the validation study carried out. The values above the lines are the estimates of the regression coefficients of the common factors and of the specific factors, that is, the contribution of each item to the factors, which is also called factor loadings. The R^2 is the variance explained by the factor for each of the items.

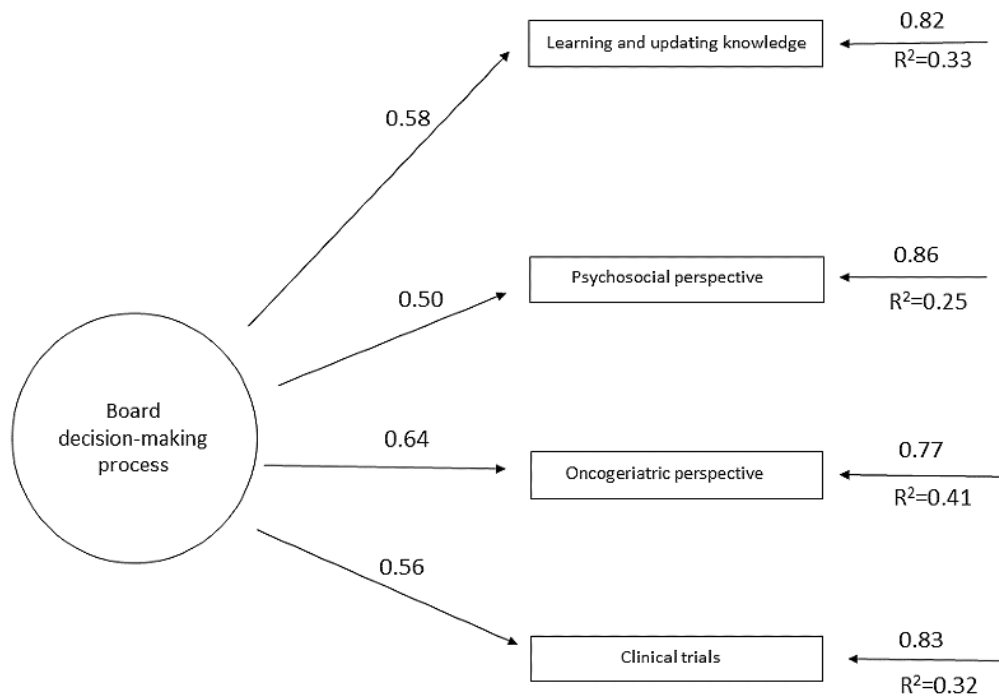


Figure 4. Optimized factor 3 model based on the confirmatory factor analysis of the validation study carried out. The values above the lines are the estimates of the regression coefficients of the common factors and of the specific factors, that is, the contribution of each item to the factors, which is also called factor loadings. The R^2 is the variance explained by the factor for each of the items.

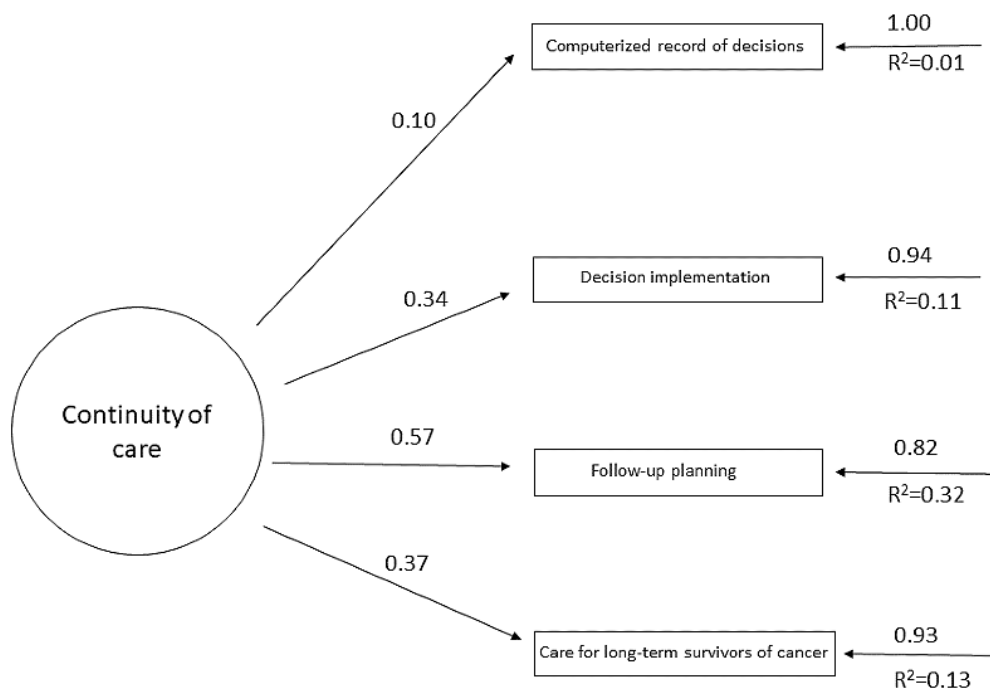


Figure 5. Optimized factor 4 model based on the confirmatory factor analysis of the validation study carried out. The values above the lines are the estimates of the regression coefficients of the common factors and of the specific factors, that is, the contribution of each item to the factors, which is also called factor loadings. The R^2 is the variance explained by the factor for each of the items.

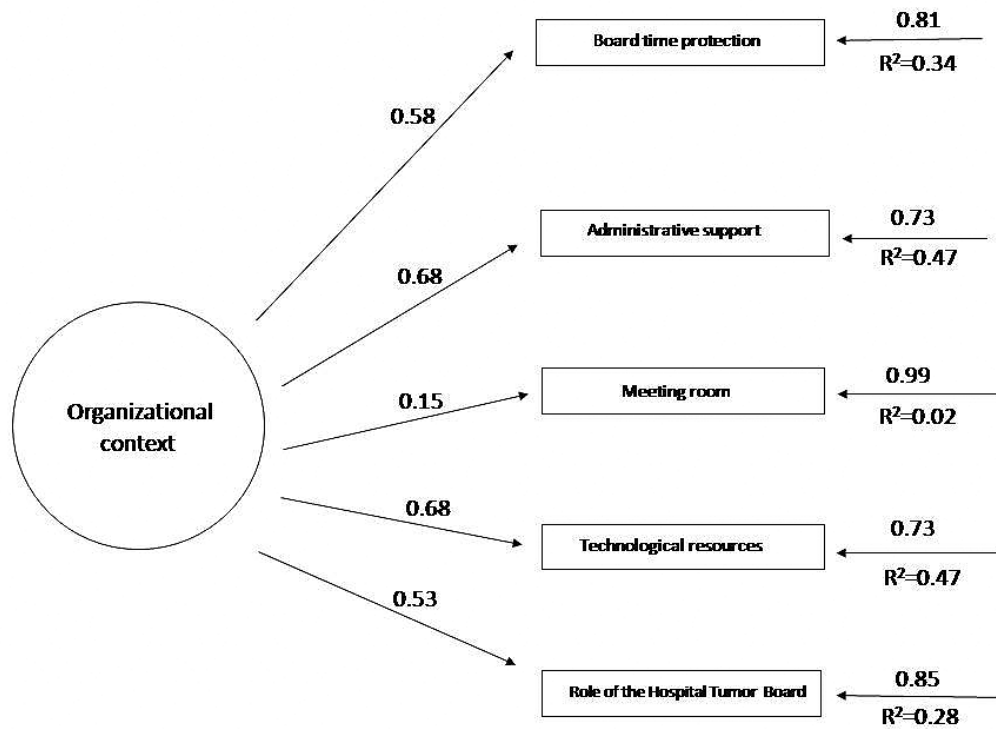


Figure 6. Optimized factor 5 model based on the confirmatory factor analysis of the validation study carried out. The values above the lines are the estimates of the regression coefficients of the common factors and of the specific factors, that is, the contribution of each item to the factors, which is also called factor loadings. The R^2 is the variance explained by the factor for each of the items.

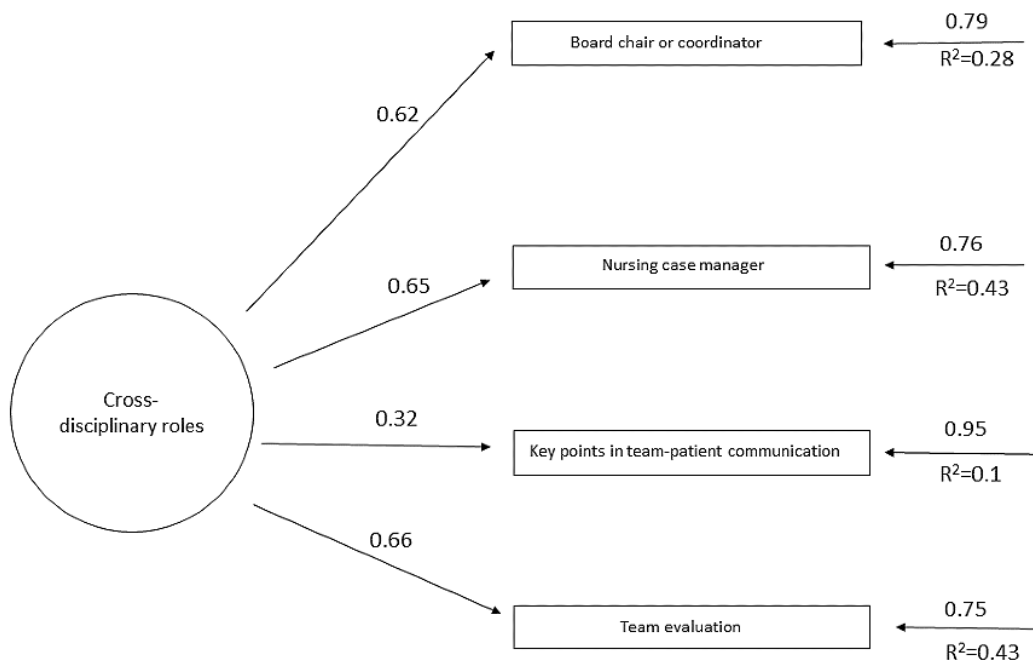


Table 3. Results of the responses to the satisfaction survey (part 1).

Item	Median score	Minimum score	Maximum score
Ease of answering the questions	8	3	10
Coherence between content of questions and reality of tumor boards	8	3	10
Adequacy of the 3-option response scale	8	3	10
Ease of creating an improvement plan in the web-based app	7	3	10
Overall satisfaction with the self-assessment	8	2	10

Table 4. Results of the responses to the satisfaction survey (part 2).

Component	Adequate (N=40), n (%)	Not adequate (N=40), n (%)
Critical components		
Attendance and representation	40 (100)	0 (0)
Patient schedule	39 (98)	1 (3)
Learning and updating knowledge	39 (98)	1 (3)
Decision implementation	38 (95)	2 (5)
Follow-up planning	35 (88)	5 (12)
Board time protection	37 (93)	3 (7)
Board chair or coordinator	39 (98)	1 (2)
Nursing case manager	23 (59)	16 (41)
Semicritical components		
Cases discussed	34 (95)	6 (5)
Patient information process	26 (73)	14 (28)
Computerized record of decisions	33 (92)	7 (8)
Key points in team-patient communication	27 (77)	13 (23)

Discussion

Overview of This Study

The AEMAC program is a web-based tool designed to facilitate the self-assessment of multidisciplinary cancer care teams with a view of improving their internal organization and scope of care. By analyzing the internal and external processes that frame the work of MDTs, we were able to identify 27 critical components that are implemented to a varying degree in the context of the SNS, as well as establish a range of possibilities for the intervention. When the AEMAC program was tested by 15 MDTs, it obtained good results on the content, acceptability, and time required, and showed reasonable psychometric properties, including internal consistency and item discrimination. The perspective acquired during the research process, which included content validation by experts, and the consistency of the results obtained in the pilot study show that the AEMAC program could be implemented in a real context.

Framework of This Study

In accordance with other authors, we consider that tumor boards stimulate the knowledge and development of the professionals involved in the care of patients with cancer, thereby encouraging teamwork and improving the quality of health care [33,34]. In this regard, several government-produced documents defining

MDTs such as the principles of multidisciplinary care in Australia [23], multidisciplinary cancer conferences in the Canadian province of Ontario [22], and the characteristics of an effective MDT in the United Kingdom [17] highlight the importance of internal cohesion and positive leadership styles. The desired consensus of a tumor board should not mask the valuable technical perspectives of its members nor should it be vulnerable to what has been called ego-based medicine [35].

The rationale and health care objectives of the AEMAC program multidisciplinary care model are based on the principles and priorities set out in the reference document *Policy Statement on multidisciplinary cancer care* [21]. For example, one component that reflects a well-developed tumor board according to the AEMAC program is that all cases treated in the hospital for a specific pathology are presented at the board meeting. While this topic “all cases discussed” has some controversy, many health professionals consider that this measure simply implies ensuring evidence-based decisions. European initiatives such as the European Reference Networks [36] share the view of securing evidence-based opinions, which is also reflected in the criteria and standards of various European cancer organizations [24-27]. These bodies are raising awareness in the SNS of the need to have well-structured and well-equipped MDTs that are answerable to a national health authority. The available evidence suggests that significant variability exists

between MDTs in terms of objectives, roles, organizational implications, performance, and access [37,38]. In the context of the SNS, these differences are accentuated by considerable decentralization in health care [13].

Comparisons With Other Instruments and Tools

The AEMAC program follows the inspiring principles of instruments such as the MDT-FIT [15]. In this case, the 3-stage process to perform the self-assessment of the MDTs and the availability of a web-based tool to perform the process have been elements that the AEMAC program has taken into account in the development of the self-assessment process, wherein the time needed to perform the whole process was 10-12 weeks, while the AEMAC program considers that the whole process can be performed in the same morning, leaving a sufficient time interval to perform the process of implementing the Improvement Plan (approximately 6-12 months). MDT-OARS [16,17], which is supposed to be a development and validation model to be followed in the case of the AEMAC program, consists of 47 items and a 5-point scale and is an instrument that each tumor committee must apply. In the case of MDT-QuIC [18], the instrument applies more like a checklist, and in its development study, the attitudes of the people who used it were assessed. The AEMAC program followed this model to assess the experiences of the people who used it. MDT-MODE [19] is based on an observational evaluation tool, where 2 professionals evaluate the behavior of the MDT. The AEMAC program is based on the individual evaluation of each of the committee members in the first phase and then sharing of the results achieved and the areas of top priority.

Strengths and Weaknesses

Strengths of the AEMAC Program

Implementing the AEMAC program could benefit clinicians, managers, and patients. For clinicians, the program not only enables comparison of tangible aspects such as technological resources (eg, use of videoconferencing) but also takes the human factor into account. Elements related to communication style or cooperation (eg, trust, implementation of decisions) are critical but tend to be ignored in evaluations or accreditation systems. For managers at all levels of the SNS, the AEMAC program offers guidance on implementing and developing an effective organizational ecosystem. In Spain, MDTs are not incorporated into the care pathway through mandatory criteria as part of health policy, as they are in the United Kingdom. The AEMAC program facilitates assessment of MDT uniformity in this context. Finally, although the defined model is for professional use, patients stand to benefit to the extent that the quality criteria include the consideration of oncogeriatric and psychosocial aspects, personalization of the information delivered to patients, and proactive organization of patients' agendas.

Weaknesses of the AEMAC Program

Although most evaluations of the content and functionality of the AEMAC program were positive, 80% of the professionals who were asked to evaluate the tool were unable to because of scheduling or other problems. We did not manage to obtain this information subsequently, although we contacted these team members individually after the on-site visits. The main feedback on the experience with the tool was provided by the people who had responsibility for the tumor board (chair and secretary). Given the uniqueness of the AEMAC questionnaire, it was validated through confirmatory factor analysis. Similarly, the difficulty of accessing teams and activating self-assessment processes during the pilot test—mainly due to the gatekeepers being informal contacts or the health professionals showing inconsistent interest in participating—meant that we were unable to include a larger sample of tumor boards. Another limitation was that the boards of only 4 hospitals participated in the pilot tests. These hospitals were similar to a majority of Spanish hospitals: accreditation for teaching, residents, students, and number of beds. As a result, adaptability issues may arise if the program is rolled out in a real context, as the AEMAC program is designed for any MDT of the SNS, and the research process and piloting covered only a few regions of Spain. If the AEMAC program is to become a reference tool, it must be promoted not only by scientific societies but also by the health authorities of each region.

Future Developments

As MDT work dynamics and structures evolve, future evaluations will determine which elements should be replaced. The AEMAC program is a dynamic quality model based on the improvement cycle that must periodically adapt to reality. The elements incorporated in the future will follow the same principle.

Conclusion

The AEMAC program is the first web-based quality self-assessment tool for evaluating MDTs in the SNS. The results obtained during the research and piloting period suggest that it could be implemented in a real context. Consensus and multidisciplinary work undoubtedly contribute to clinical effectiveness insofar as professionals specialized in tumor pathologies (eg, esophageal cancer), organ systems (eg, gynecological tumors), or patient profiles (eg, pediatric oncology) are responsible for clinical coordination and communication with patients and families throughout all stages of the disease. To this end, the AEMAC program can provide a comprehensive reflection on the organizational, technological, and cultural elements that must be taken into account to improve the care received by patients with cancer.

Acknowledgments

The Autoevaluación de Equipos Multidisciplinares de Atención al Cáncer (AEMAC) program has benefited from the involvement of the most relevant scientific societies in cancer care in Spain: Spanish Association of Surgeons, Spanish Society of Medical Oncology, Spanish Society of Radiation Oncology, Spanish Society of Pathological Anatomy, Spanish Society of Medical

Radiology, Spanish Society of Oncology Nursing, Spanish Society of Pediatric Hematology and Oncology, Spanish Society of Hospital Pharmacy, and Spanish Society of Healthcare Quality. With their assistance, we were able to validate the tool, identify experts, and find contacts for the pilot study. In addition, all the societies promoted the use of the AEMAC program by publishing on their respective websites the agreement on the development of multidisciplinary cancer care teams in the Spanish National Health System and the use of the AEMAC program. The authors would like to thank all the hospitals that participated in the development and validation of the AEMAC program and would like to thank the members of PROMETEO 173 Research Group (Generalitat Valenciana, 2017) for their assistance. This work was supported by nonrestrictive funding from Merck Sharp & Dohme, Spain. No staff from Merck Sharp & Dohme participated in the design, execution, data capture, interpretation of results, or elaboration of the conclusions of this study.

Authors' Contributions

JP and JMB were in charge of the design of the Autoevaluación de Equipos Multidisciplinares de Atención al Cáncer (AEMAC) Program and all the theoretical underpinning. MG and JJM carried out all the pilot tests of the tool in the 4 hospitals. IM, LF, and JAG acted as leaders in the collection of information from the different tumor committees for the validation process of the tool. All members of the AEMAC program research team actively participated in the piloting of the tool in each of the phases. JP, MG, JMB, and JJM drafted and reviewed the final manuscript.

The AEMAC program research team consisted of Gemma Mancebo (Servei de Ginecologia i Obstetrícia, Coordinadora de la Unitat Funcional de Càncer Ginecològic, Hospital del Mar, Barcelona), Ruth Vera (Servicio de Oncología Médica, Complejo Hospitalario de Navarra, Pamplona), Ignacio Inchaurreaga Álvarez (Servicio de Oncología Médica, Hospital Universitario Dr. Peset), Segundo Gómez Abril (Servicio de Oncología Médica, Hospital Universitario Dr. Peset, Valencia), Margarita Cueto Callejón (Servicio de Ginecología y Obstetrícia, Hospital Universitario de Fuenlabrada, Fuenlabrada), Belén Pérez-Mies (Servicio Anatomía patológica Hospital Universitario Ramón y Cajal, Facultad de Medicina de Alcalá de Henares, Instituto Ramón y Cajal para la investigación y CIBERONC, Instituto Carlos III), Gonzalo Varela (Profesor titular de cirugía torácica de la Universidad de Salamanca), and Marta Moro Agud (Subdirección médica, Hospital Universitario La Paz).

Conflicts of Interest

Merck Sharp & Dohme, Spain financed the design and management costs of the website where the responses were recorded. During this study, JJM and MG were supported by a grant from the Generalitat Valenciana (reference Prometeo 2017/0173).

Multimedia Appendix 1

Supplementary data.

[[DOCX File , 38 KB - jmir_v24i3e29063_app1.docx](#)]

Multimedia Appendix 2

Phase 1: Individual self-assessment.

[[PNG File , 480 KB - jmir_v24i3e29063_app2.png](#)]

Multimedia Appendix 3

Phase 2: Group prioritization process.

[[PNG File , 457 KB - jmir_v24i3e29063_app3.png](#)]

Multimedia Appendix 4

Phase 3: Creating an improvement plan.

[[PNG File , 258 KB - jmir_v24i3e29063_app4.png](#)]

Multimedia Appendix 5

Experience survey to assess critical issues of the Autoevaluación de Equipos Multidisciplinares de Atención al Cáncer program.

[[DOCX File , 90 KB - jmir_v24i3e29063_app5.docx](#)]

References

1. Atun R, Martin-Moreno J, Ogawa T. Analysis of National Cancer Control Programmes in Europe. 2009. URL: <https://core.ac.uk/download/pdf/297067.pdf> [accessed 2021-10-04]
2. El Saghir NS, Keating NL, Carlson RW, Khoury KE, Fallowfield L. Tumor boards: optimizing the structure and improving efficiency of multidisciplinary management of patients with cancer worldwide. *Am Soc Clin Oncol Educ Book* 2014:e461-e466 [FREE Full text] [doi: [10.14694/EdBook_AM.2014.34.e461](https://doi.org/10.14694/EdBook_AM.2014.34.e461)] [Medline: [24857140](https://pubmed.ncbi.nlm.nih.gov/24857140/)]
3. Brar SS, Hong NL, Wright FC. Multidisciplinary cancer care: does it improve outcomes? *J Surg Oncol* 2014 Oct;110(5):494-499. [doi: [10.1002/jso.23700](https://doi.org/10.1002/jso.23700)] [Medline: [24986235](https://pubmed.ncbi.nlm.nih.gov/24986235/)]

4. Hong NJL, Wright FC, Gagliardi AR, Paszat LF. Examining the potential relationship between multidisciplinary cancer care and patient survival: an international literature review. *J Surg Oncol* 2010 Aug 01;102(2):125-134. [doi: [10.1002/jso.21589](https://doi.org/10.1002/jso.21589)] [Medline: [20648582](https://pubmed.ncbi.nlm.nih.gov/20648582/)]
5. Denton E, Conron M. Improving outcomes in lung cancer: the value of the multidisciplinary health care team. *J Multidiscip Healthc* 2016;9:137-144 [FREE Full text] [doi: [10.2147/JMDH.S76762](https://doi.org/10.2147/JMDH.S76762)] [Medline: [27099511](https://pubmed.ncbi.nlm.nih.gov/27099511/)]
6. Foster TJ, Bouchard-Fortier A, Olivotto IA, Quan ML. Effect of Multidisciplinary Case Conferences on Physician Decision Making: Breast Diagnostic Rounds. *Cureus* 2016 Nov 24;8(11):e895 [FREE Full text] [doi: [10.7759/cureus.895](https://doi.org/10.7759/cureus.895)] [Medline: [28018765](https://pubmed.ncbi.nlm.nih.gov/28018765/)]
7. Pillay B, Wootten AC, Crowe H, Corcoran N, Tran B, Bowden P, et al. The impact of multidisciplinary team meetings on patient assessment, management and outcomes in oncology settings: A systematic review of the literature. *Cancer Treat Rev* 2016 Jan;42:56-72. [doi: [10.1016/j.ctrv.2015.11.007](https://doi.org/10.1016/j.ctrv.2015.11.007)] [Medline: [26643552](https://pubmed.ncbi.nlm.nih.gov/26643552/)]
8. Charara RN, Kreidieh FY, Farhat RA, Al-Feghali KA, Khoury KE, Haydar A, et al. Practice and Impact of Multidisciplinary Tumor Boards on Patient Management: A Prospective Study. *J Glob Oncol* 2017 Jun;3(3):242-249 [FREE Full text] [doi: [10.1200/JGO.2016.004960](https://doi.org/10.1200/JGO.2016.004960)] [Medline: [28717766](https://pubmed.ncbi.nlm.nih.gov/28717766/)]
9. Prades J, Borrás JM. Shifting sands: adapting the multidisciplinary team model to technological and organizational innovations in cancer care. *Future Oncol* 2014 Oct;10(13):1995-1998 [FREE Full text] [doi: [10.2217/fon.14.125](https://doi.org/10.2217/fon.14.125)] [Medline: [25396771](https://pubmed.ncbi.nlm.nih.gov/25396771/)]
10. Prades J, Remue E, van Hoof E, Borrás JM. Is it worth reorganising cancer services on the basis of multidisciplinary teams (MDTs)? A systematic review of the objectives and organisation of MDTs and their impact on patient outcomes. *Health Policy* 2015 Apr;119(4):464-474. [doi: [10.1016/j.healthpol.2014.09.006](https://doi.org/10.1016/j.healthpol.2014.09.006)] [Medline: [25271171](https://pubmed.ncbi.nlm.nih.gov/25271171/)]
11. Catt S, Fallowfield L, Jenkins V, Langridge C, Cox A. The informational roles and psychological health of members of 10 oncology multidisciplinary teams in the UK. *Br J Cancer* 2005 Nov 14;93(10):1092-1097 [FREE Full text] [doi: [10.1038/sj.bjc.6602816](https://doi.org/10.1038/sj.bjc.6602816)] [Medline: [16234824](https://pubmed.ncbi.nlm.nih.gov/16234824/)]
12. Ministerio DS, Servicios SEI. Cancer strategy of the Spanish national health system 2009. Ministerio de Sanidad, Servicios Sociales e Igualdad Centro de Publicaciones Madrid. 2009. URL: https://www.mscbs.gob.es/organizacion/sns/planCalidadSNS/pdf/Cancer_Strategy_of_the_Spanish_2009.pdf [accessed 2021-03-19]
13. Prades J, Borrás JM. Multidisciplinary cancer care in Spain, or when the function creates the organ: qualitative interview study. *BMC Public Health* 2011 Feb 28;11:141 [FREE Full text] [doi: [10.1186/1471-2458-11-141](https://doi.org/10.1186/1471-2458-11-141)] [Medline: [21356063](https://pubmed.ncbi.nlm.nih.gov/21356063/)]
14. Berlanga P, Segura V, Juan Ribelles A, Sánchez de Toledo P, Acha T, Castel V, et al. Paediatric tumour boards in Spain: a national survey. *Clin Transl Oncol* 2016 Sep;18(9):931-936. [doi: [10.1007/s12094-015-1466-9](https://doi.org/10.1007/s12094-015-1466-9)] [Medline: [26693730](https://pubmed.ncbi.nlm.nih.gov/26693730/)]
15. What is MDT-FIT? MDT-FIT. URL: <https://www.mdffit.co.uk> [accessed 2017-06-12]
16. Taylor C, Brown K, Lamb B, Harris J, Sevdalis N, Green JSA. Developing and testing TEAM (Team Evaluation and Assessment Measure), a self-assessment tool to improve cancer multidisciplinary teamwork. *Ann Surg Oncol* 2012 Dec;19(13):4019-4027. [doi: [10.1245/s10434-012-2493-1](https://doi.org/10.1245/s10434-012-2493-1)] [Medline: [22820934](https://pubmed.ncbi.nlm.nih.gov/22820934/)]
17. Taylor C, Atkins L, Richardson A, Tarrant R, Ramirez A. Measuring the quality of MDT working: an observational approach. *BMC Cancer* 2012 May 29;12:202 [FREE Full text] [doi: [10.1186/1471-2407-12-202](https://doi.org/10.1186/1471-2407-12-202)] [Medline: [22642614](https://pubmed.ncbi.nlm.nih.gov/22642614/)]
18. Lamb BW, Sevdalis N, Vincent C, Green JSA. Development and evaluation of a checklist to support decision making in cancer multidisciplinary team meetings: MDT-QuIC. *Ann Surg Oncol* 2012 Jun;19(6):1759-1765. [doi: [10.1245/s10434-011-2187-0](https://doi.org/10.1245/s10434-011-2187-0)] [Medline: [22207050](https://pubmed.ncbi.nlm.nih.gov/22207050/)]
19. Lamb BW, Wong HWL, Vincent C, Green JSA, Sevdalis N. Teamwork and team performance in multidisciplinary cancer teams: development and evaluation of an observational assessment tool. *BMJ Qual Saf* 2011 Oct;20(10):849-856. [doi: [10.1136/bmjqs.2010.048660](https://doi.org/10.1136/bmjqs.2010.048660)] [Medline: [21610266](https://pubmed.ncbi.nlm.nih.gov/21610266/)]
20. Soukup T, Lamb BW, Arora S, Darzi A, Sevdalis N, Green JS. Successful strategies in implementing a multidisciplinary team working in the care of patients with cancer: an overview and synthesis of the available literature. *J Multidiscip Healthc* 2018;11:49-61 [FREE Full text] [doi: [10.2147/JMDH.S117945](https://doi.org/10.2147/JMDH.S117945)] [Medline: [29403284](https://pubmed.ncbi.nlm.nih.gov/29403284/)]
21. European Partnership Action Against Cancer consensus group, Borrás JM, Albrecht T, Audisio R, Briens E, Casali P, et al. Policy statement on multidisciplinary cancer care. *Eur J Cancer* 2014 Feb;50(3):475-480. [doi: [10.1016/j.ejca.2013.11.012](https://doi.org/10.1016/j.ejca.2013.11.012)] [Medline: [24321260](https://pubmed.ncbi.nlm.nih.gov/24321260/)]
22. Wright FC, De Vito C, Langer B, Hunter A, Expert Panel on Multidisciplinary Cancer Conference Standards. Multidisciplinary cancer conferences: a systematic review and development of practice standards. *Eur J Cancer* 2007 Apr;43(6):1002-1010. [doi: [10.1016/j.ejca.2007.01.025](https://doi.org/10.1016/j.ejca.2007.01.025)] [Medline: [17329094](https://pubmed.ncbi.nlm.nih.gov/17329094/)]
23. Principles of multidisciplinary care. Australia Government. URL: <https://canceraustralia.gov.au/clinical-best-practice/multidisciplinary-care/all-about-multidisciplinary-care/principles-multidisciplinary-care> [accessed 2020-02-19]
24. Andritsch E, Beishon M, Bielack S, Bonvalot S, Casali P, Crul M, et al. ECCO Essential Requirements for Quality Cancer Care: Soft Tissue Sarcoma in Adults and Bone Sarcoma. A critical review. *Crit Rev Oncol Hematol* 2017 Feb;110:94-105 [FREE Full text] [doi: [10.1016/j.critrevonc.2016.12.002](https://doi.org/10.1016/j.critrevonc.2016.12.002)] [Medline: [28109409](https://pubmed.ncbi.nlm.nih.gov/28109409/)]
25. Beets G, Sebag-Montefiore D, Andritsch E, Arnold D, Beishon M, Crul M, et al. ECCO Essential Requirements for Quality Cancer Care: Colorectal Cancer. A critical review. *Crit Rev Oncol Hematol* 2017 Feb;110:81-93 [FREE Full text] [doi: [10.1016/j.critrevonc.2016.12.001](https://doi.org/10.1016/j.critrevonc.2016.12.001)] [Medline: [28109408](https://pubmed.ncbi.nlm.nih.gov/28109408/)]

26. Wilson ARM, Marotti L, Bianchi S, Biganzoli L, Claassen S, Decker T, EUSOMA (European Society of Breast Cancer Specialists). The requirements of a specialist Breast Centre. *Eur J Cancer* 2013 Nov;49(17):3579-3587. [doi: [10.1016/j.ejca.2013.07.017](https://doi.org/10.1016/j.ejca.2013.07.017)] [Medline: [23968730](https://pubmed.ncbi.nlm.nih.gov/23968730/)]
27. Accreditation and designation user manual ver 3.0. OEI Standards. URL: https://www.oeci.eu/Accreditation/Attachments/OEAI_A&D_MANUAL_3.0.2019_DEF.pdf [accessed 2021-03-24]
28. Yin R. *Case Study Research: Design and Methods (Applied Social Research Methods)*. London: Sage Publications; 2009:93-95.
29. Flyvbjerg B. Case study. In: *The Sage Handbook of Qualitative Research*, 4th ed. Thousand Oaks, CA: Sage Publications; 2011:311-316.
30. AEMAC Program. URL: <https://www.programaemac.org/> [accessed 2021-10-04]
31. Tabachnick B, Fidell L, Ullman J. *Using multivariate statistics*. Pearson. URL: <https://www.pearsonhighered.com/assets/preface/0/1/3/4/0134790545.pdf> [accessed 2021-08-01]
32. Hu L, Bentler P. Structural Equation Modeling: A Multidisciplinary Journal Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Multidiscip J Internet* 2009 Nov 03:1-55 [FREE Full text] [doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)]
33. Jazieh AR. Tumor boards: beyond the patient care conference. *J Cancer Educ* 2011 Sep;26(3):405-408. [doi: [10.1007/s13187-011-0209-z](https://doi.org/10.1007/s13187-011-0209-z)] [Medline: [21380508](https://pubmed.ncbi.nlm.nih.gov/21380508/)]
34. Pérez-Jover V, Mira JJ, Salas D, Ibañez J, Miranda J, Guilabert M, et al. [Results assessment of the Valencian Community Cancer Plan: the point of view of patients, relatives, citizens and professionals]. *Rev Calid Asist* 2013;28(2):71-83. [doi: [10.1016/j.cali.2012.10.001](https://doi.org/10.1016/j.cali.2012.10.001)] [Medline: [23201186](https://pubmed.ncbi.nlm.nih.gov/23201186/)]
35. Yeh J, Kagan A, Steckel R. Ego-based medicine. *Int J Radiat Oncol Biol Phys* 2013 May 01;86(1):13. [doi: [10.1016/j.ijrobp.2012.11.029](https://doi.org/10.1016/j.ijrobp.2012.11.029)] [Medline: [23395432](https://pubmed.ncbi.nlm.nih.gov/23395432/)]
36. European reference networks. European Commission. URL: https://ec.europa.eu/health/ern_en [accessed 2018-09-20]
37. Saini KS, Taylor C, Ramirez A, Palmieri C, Gunnarsson U, Schmoll HJ, et al. Role of the multidisciplinary team in breast cancer management: results from a large international survey involving 39 countries. *Ann Oncol* 2012 Apr;23(4):853-859 [FREE Full text] [doi: [10.1093/annonc/mdr352](https://doi.org/10.1093/annonc/mdr352)] [Medline: [21821551](https://pubmed.ncbi.nlm.nih.gov/21821551/)]
38. Ruhstaller T, Roe H, Thürlimann B, Nicoll JJ. The multidisciplinary meeting: An indispensable aid to communication between different specialties. *Eur J Cancer* 2006 Oct;42(15):2459-2462. [doi: [10.1016/j.ejca.2006.03.034](https://doi.org/10.1016/j.ejca.2006.03.034)] [Medline: [16934974](https://pubmed.ncbi.nlm.nih.gov/16934974/)]

Abbreviations

AEMAC: Autoevaluación de Equipos Multidisciplinares de Atención al Cáncer

MDT: multidisciplinary team

MDT-FIT: multidisciplinary team-feedback for improving teamworking

MDT-MODE: multidisciplinary team-metric of decision making

MDT-OARS: multidisciplinary team-observational assessment rating scale

MDT-QuIC: multidisciplinary team-quality improvement checklist

SNS: Spanish National Health System

Edited by R Kukafka; submitted 24.03.21; peer-reviewed by S Rostam Niakan Kalhori, G Maifredi; comments to author 13.05.21; revised version received 02.08.21; accepted 12.08.21; published 10.03.22.

Please cite as:

Guilabert M, Prades J, Borrás JM, Maestu I, Guerra JA, Fumadó L, Mira JJ, AEMAC Program Research Team

A Web-Based Self-assessment Model for Evaluating Multidisciplinary Cancer Teams in Spain: Development and Validation Pilot Study

J Med Internet Res 2022;24(3):e29063

URL: <https://www.jmir.org/2022/3/e29063>

doi: [10.2196/29063](https://doi.org/10.2196/29063)

PMID: [35266870](https://pubmed.ncbi.nlm.nih.gov/35266870/)

©Mercedes Guilabert, Joan Prades, Josep M Borrás, Inmaculada Maestu, Juan Antonio Guerra, Lluís Fumadó, José Joaquín Mira, AEMAC Program Research Team. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org/>), 10.03.2022. 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.

Original Paper

Contextualizing Engagement With Health Information on Facebook: Using the Social Media Content and Context Elicitation Method

Yonaira M Rivera^{1,2}, MPH, PhD; Meghan B Moran³, PhD; Johannes Thrul⁴, MS, PhD; Corinne Joshu⁵, PhD; Katherine C Smith³, PhD

¹Department of Communication, School of Communication & Information, Rutgers University, New Brunswick, NJ, United States

²Cancer Prevention and Control Program, Rutgers Cancer Institute of New Jersey, New Brunswick, NJ, United States

³Department of Health, Behavior & Society, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

⁴Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

⁵Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, United States

Corresponding Author:

Yonaira M Rivera, MPH, PhD

Department of Communication

School of Communication & Information

Rutgers University

4 Huntington St

New Brunswick, NJ, 08901

United States

Phone: 1 (848) 932 7612

Email: yonaira.rivera@rutgers.edu

Abstract

Background: Most of what is known regarding health information engagement on social media stems from quantitative methodologies. Public health literature often quantifies engagement by measuring likes, comments, and/or shares of posts within health organizations' Facebook pages. However, this content may not represent the health information (and misinformation) generally available to and consumed by platform users. Furthermore, some individuals may prefer to engage with information without leaving quantifiable digital traces. Mixed methods approaches may provide a way of surpassing the constraints of assessing engagement with health information by using only currently available social media metrics.

Objective: This study aims to discuss the limitations of current approaches in assessing health information engagement on Facebook and presents the social media content and context elicitation method, a qualitatively driven, mixed methods approach to understanding engagement with health information and how engagement may lead to subsequent actions.

Methods: Data collection, management, and analysis using the social media content and context elicitation method are presented. This method was developed for a broader study exploring how and why US Latinos and Latinas engage with cancer prevention and screening information on Facebook. The study included 20 participants aged between 40 and 75 years without cancer who participated in semistructured, in-depth interviews to discuss their Facebook use and engagement with cancer information on the platform. Participants accessed their Facebook account alongside the researcher, typed *cancer* in the search bar, and discussed cancer-related posts they engaged with during the previous 12 months. Engagement was defined as liking, commenting, and/or sharing a post; clicking on a post link; reading an article in a post; and/or watching a video within a post. Content engagement prompted questions regarding the reasons for engagement and whether engagement triggered further action. Data were managed using MAXQDA (VERBI GmbH) and analyzed using thematic and content analyses.

Results: Data emerging from the social media content and context elicitation method demonstrated that participants mainly engaged with cancer prevention and screening information by viewing and/or reading content (48/66, 73%) without liking, commenting, or sharing it. This method provided rich content regarding how US Latinos and Latinas engage with and act upon cancer prevention and screening information on Facebook. We present 2 emblematic cases from the main study to exemplify the additional information and context elicited from this methodology, which is currently lacking from quantitative approaches.

Conclusions: The social media content and context elicitation method allows a better representation and deeper contextualization of how people engage with and act upon health information and misinformation encountered on social media. This method may be applied to future studies regarding how to best communicate health information on social media, including how these affect assessments of message credibility and accuracy, which can influence health outcomes.

(*J Med Internet Res* 2022;24(3):e25243) doi:[10.2196/25243](https://doi.org/10.2196/25243)

KEYWORDS

mixed methods; data collection; social media; cancer; health information; Facebook; digital health

Introduction

Background

The rise of health misinformation in today's social media landscape has prompted a need to better understand how and why individuals engage with this content, as well as its ramifications on health outcomes. Although this topic has gained notoriety in light of the COVID-19 pandemic and its accompanying infodemic, calls for research addressing health misinformation and its unique impact on underserved populations have been present since late 2018 [1]. These calls acknowledge that in addition to defining the prevalence and trends of health misinformation, researchers need to develop approaches that better understand the context of misinformation exchange on social media, the intra- and interpersonal dynamics that influence engagement with content, and how health consequences may stem from these interactions [1].

Reaching populations with evidence-based content through social media has become an important effort to counteract the spread of health misinformation [2,3]. If leveraged correctly, these platforms can be used to encourage participatory communication by fostering user engagement via posts, pictures, videos, and other forms of information sharing [4]. This conceptualization of social media as participatory frames engagement as a way for health organizations to communicate with audiences directly [5] and is typically assessed by evaluating how users respond to content posted on the platform. By playing an active role in conversations about health topics, organizations can also ensure that trust and credibility are established through the dissemination of accurate information [5].

Quantitative methods have undoubtedly helped identify health misinformation trends on social media [6-9]. However, these data are increasingly difficult to obtain [10], do not provide important contextual information regarding what motivates engagement and dissemination among vulnerable populations with poor health outcomes, and cannot capture the effects of misinformation on behavior. Mixed methods approaches that explore the role of these components in the spread of misinformation are necessary to design interventions that minimize and halt dissemination. Mixed methods research comprehensively and purposefully uses both qualitative and quantitative techniques to address an overarching research question that cannot be fully explored and contextualized by either method independently [11]. As such, this paper presents the social media content and context elicitation method, which is a novel approach that incorporates qualitative methods to better contextualize engagement with health information on

social media and how this may lead to subsequent actions. This paper first discusses the limitations of the current operationalizations of engagement with health information on social media. This is followed by a detailed description of the social media content and context elicitation method, which was developed to obtain survey data, interviews, and computer screen recordings of cancer-related posts on Latino and Latina participants' Facebook accounts for quantitative and qualitative analysis. Then, 2 case studies are presented to exemplify the additional information elicited from this methodology, which is currently lacking from other approaches. Finally, we discuss how incorporating qualitative methods, such as those outlined in this paper, allows a better representation of how people engage with health information in reality and provides insights for researchers interested in this type of work.

Assessing Engagement With Health Information on Facebook

Facebook is among the most popular social media platforms worldwide, with >2.3 billion active users [12]. Second in popularity only to YouTube, 74% of US Facebook users visit the platform on a daily basis [13]. Entertainment, social interaction, and passing time are among the reasons individuals report using Facebook [14]. Facebook has also been a source of health information and social support [15], making it a useful place to engage with general audiences about health topics. Many public health organizations have established a presence on Facebook by creating a Facebook page, which provides a space for businesses and organizations to publicly share information with platform users. Facebook pages provide a direct way for these organizations to deliver evidence-based health information to Facebook users, which is of paramount importance in a social media environment with increasingly unreliable information [3]. Facebook page administrators also have the ability to monitor social media metrics, providing a way for health organizations to operationalize audience engagement with posted content.

Assessing engagement with health-related information on social media is of particular importance as it is a precursor to multiple outcomes, such as increased awareness, knowledge, and behavior change [16,17]. Most studies have assessed engagement by collecting and analyzing data on the likes, comments, and/or shares of posts within an organization's Facebook page [18-24]. For example, Strekalova and Krieger [24] reported that cancer-related posts on the National Cancer Institute's Facebook page had a significantly higher number of likes, comments, and shares when they contained images (vs videos, embedded links, or text). Similarly, Srivastava et al [18] found that posts on the American Cancer Society's Facebook

page were more likely to be liked or shared when they contained images or videos, whereas text-based posts were more likely to elicit user comments. Meanwhile, Klippert and Schaper [25] expanded their definition of engagement by including metrics for post reach and clicks on embedded links—both of which are also available to Facebook page administrators. Finally, other studies have captured engagement with cancer information publicly available on Facebook [19,20,26] or Facebook groups [27-30]. Facebook groups differ from Facebook pages in that they can be public or private but do not offer detailed social media metrics and audience insights (although group administrators may extract raw data for analysis through Facebook's application programming interface). In such cases, engagement has been assessed by quantifying likes, comments, and shares, as these metrics are visible to anyone with access to the posted content.

Limitations of Quantitative Assessments of Engagement

Measuring engagement with health-related content through these metrics is useful for organizations wanting to assess the success of a social media campaign. It can also provide insight into message factors that may enhance engagement with health information on social media [31]. However, the existing metrics have important limitations. On Facebook, one of these limitations relates to how users are exposed to content. In order for a post from a Facebook page to appear on a person's news feed, a person must either follow the page or have a Facebook friend who engages with a post from the page. Additional ways users can be exposed to health-related content from a Facebook page are through paid advertising or a Facebook video recommendation, which is based on a video's popularity or other people and pages a person follows [32]. Even then, the appearance of this content on a person's news feed is influenced by Facebook's constantly changing algorithm, which favors content that individuals engage with most often [33]. This has an impact on whether specific health information emerges on a person's news feed when they log into their Facebook account. As such, engagement with content on a health-related Facebook page may not be emblematic of how the general population engages with such information on Facebook. It is likely that many individuals following a health-related Facebook page are already interested in that particular topic. However, there are many people who may not have an active interest in health information that health organizations are trying to reach, such as healthy individuals who are the target audience for prevention and screening messages. Furthermore, focusing on measuring engagement with evidence-based content posted by health organizations does not fully capture the health information landscape on Facebook, which includes user-generated or shared health misinformation that may not come from reliable sources (eg, a COVID-19-related post dispelling misinformation about vaccine efficiency shared by a Facebook friend with no links to original sources).

Another limitation to quantifying likes, comments, and shares is that these are crude measures of engagement. Although these metrics allow researchers to quantify how some Facebook users visibly engage with health information that is publicly available or posted within a Facebook group, they exclude individuals

who do not perform these actions yet still consume health information on the platform [18,24]. Information consumption and lurking—generally defined as reading posts on the web without responding—have been seen as an active and participative form of web-based behavior [34]. Lurking may occur because of environmental, relationship, security, and individual reasons [35]. For example, the quality of a message may be poor (environmental), the user may not feel part of the web-based community (relationship) or have privacy concerns (security), or the person's needs may be satisfied by just reading a post (individual) [35].

Moreover, although newer Facebook applications, such as CrowdTangle, allow researchers to capture additional engagement metrics (such as post views) [36], these metrics are limited in only establishing general trends with content that is publicly available on the platform. Furthermore, these crude measures fail to capture if and how engagement with health information and misinformation may influence individuals to act upon this information elsewhere. Potential actions may be as small as discussing the information with a friend through messaging apps or as large as incorporating preventive behavior into one's lifestyle. Understanding these complexities inevitably requires new approaches to help contextualize the impact of engagement on health outcomes.

In response to these needs, we developed the social media content and context elicitation method. This method elicits data concurrently during one-on-one in-person encounters where the participants access their social media profile, scroll through relevant content, and contextualize content engagement with the researcher. In the following sections, we outline the process of collecting, managing, and analyzing elicited data and provide examples of the robust findings that this method provides. We hope that such detail—particularly surrounding data collection and management—enables other scholars to replicate and/or adapt these methods for related studies.

Methods

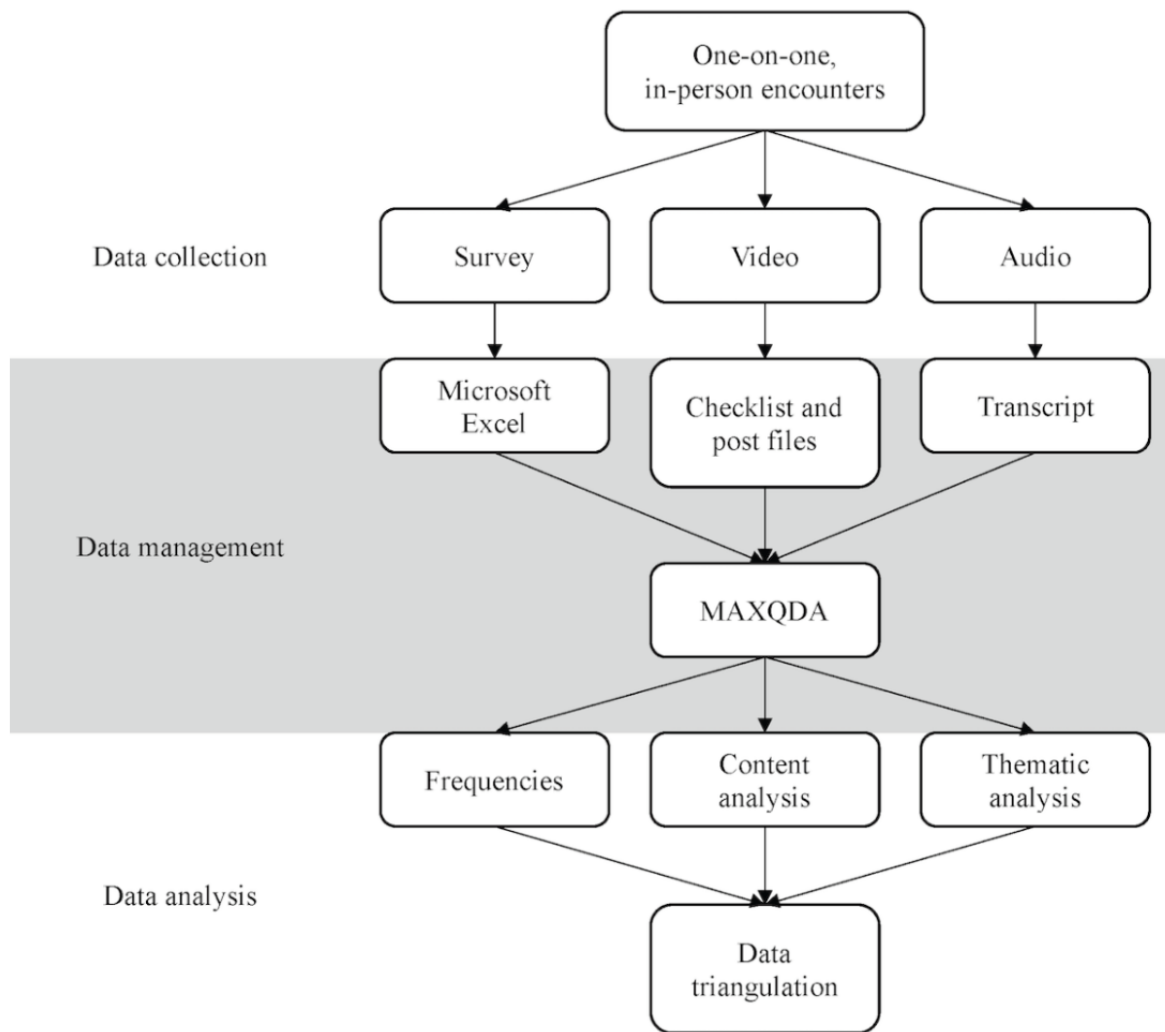
Overview

The methods discussed in this paper were developed for an exploratory, convergent parallel study assessing how and why Latino and Latina adults aged 40 to 75 years without a history of cancer engage with and act upon cancer prevention and screening information or misinformation on Facebook (published elsewhere) [37]. For this study, 20 self-identified Latinos and Latinas aged 40 to 75 years with no history of cancer participated in semistructured, in-depth interviews to discuss their Facebook use and engagement with cancer information on the platform. This diverse population not only avidly uses Facebook but also faces high cancer health disparities: cancer is the leading cause of death among US Latinos and Latinas [38], and cancer incidence rates are highest for screenable cancers linked to preventable behaviors (breast, prostate, and colorectal) [39]. Please refer to the original publication for a full description of the study and the main findings [37].

The social media content and context elicitation method developed for this study comprised three parts: (1) a short survey collecting demographics, health-related information seeking, and Facebook use data; (2) computer screen recordings of cancer posts appearing on participants' Facebook during the past 12 months; and (3) semistructured, in-depth interviews discussing Facebook use and engagement with cancer posts on Facebook

(Figure 1). Participants were recruited through flyers, word of mouth, and Facebook advertisements. Interviews were conducted in the participants' language of preference (English or Spanish) by the lead researcher, who is bilingual. All interviews were conducted during the summer of 2018 and lasted approximately 2 hours.

Figure 1. Study design using the social media content and context elicitation method to capture engagement with cancer information on Facebook. Each participant underwent all points of data collection.



Data Collection

After providing oral consent, participants completed a short survey collecting demographic variables, basic health-related information seeking, and Facebook use information. This survey provided descriptive insight into the uses and gratifications experienced by Latinos and Latinas on Facebook and other contextual factors that may affect engagement with cancer prevention and screening information on the platform. Following the survey, the researcher began the semistructured interviews, which were audio recorded in their entirety. Using the survey responses as a guide, the researcher asked participants to elaborate on their regular Facebook use patterns and interactions, the extent to which they encountered health information (including cancer information) on Facebook, and what they believed Facebook's role was in sharing information. Afterward, participants logged into their Facebook account using a private

browser on a research laptop and proceeded to turn off the Facebook Messenger feature to avoid being interrupted during the study. The researcher then documented the total number of friends, groups, and pages the participants followed, including how many of these were cancer-related groups or pages.

The participants then went to the search feature on Facebook, which allows Facebook users to search for content posted on the platform. This feature allows users to sort search results using multiple filters, such as *Sort by*, *Posted by*, and *Date posted*. For this study, participants were asked to enter the term *cancer* into the search bar. Once the search results emerged, they were filtered chronologically (*Sort by Most recent*) and by friends and groups the participant followed on Facebook (*Posted by Your friends and groups*). The resulting posts represented all posts that included the word *cancer* that could have potentially appeared on participants' news feeds when they

previously logged into Facebook and corresponded to content either posted by their friends or groups or any other publicly available posts that a friend liked or commented on. The researcher then proceeded to explain the process of jointly scrolling through the past 6 to 12 months of cancer-related posts to discuss posts they recalled seeing and engaging with. Any questions that participants had about the process were discussed before beginning.

Once the participant agreed, the researcher began recording the computer screen using QuickTime Player (version 10.4; Apple), which captures both audio and the computer screen. The researcher and participants jointly scrolled through the content to identify any posts the participants recalled having seen and whether they engaged with the post. Engagement was defined as any combination of the following: liking a post; commenting on a post; sharing a post; clicking on a post link; reading an article in a post; or watching a video within a post. If the post included any video or embedded link, participants were asked if they recalled watching the video or clicking on the link. If so, these were opened to capture the full content.

In addition to capturing the cancer posts that appeared on participants' Facebook through computer screen recordings, engagement with content prompted the researcher to use a semistructured, in-depth interview guide to ask questions regarding the reasons participants interacted with the post and whether engagement triggered further action. Examples of action included (but were not limited to) searching for additional cancer information or scheduling a cancer screening appointment. In-depth interviews were selected for this study as they allow for the exploration of new issues in depth and elaborate on individuals' thoughts and behaviors [40], an important facet in exploring how source and content characteristics influence engagement with cancer information on Facebook and any potential subsequent action. Interview guide questions were informed by the Uses and Gratification Theory [41] and the Comprehensive Model of Information Seeking [42,43]. The interview guide covered the following domains: reasons for engagement with cancer information, relationship to the cancer information source, roles of the cancer information source in delivering information on Facebook, perceptions about posted cancer information content and attributes, the ways that source credibility and content accuracy are assessed, and actions triggered by engagement with this information. In cases where participants recalled engaging with a post in ways other than liking, commenting, and/or sharing the post, the participant was asked to elaborate on this type of engagement. The researcher also collected notes regarding each post the participant either recalled or engaged with using a checklist.

Throughout the scrolling process, multiple participants had copious amounts of cancer-related information emerging in their searches, most of which were not specific to prevention and screening topics (eg, survivorship, cancer research, and fundraising). As the purpose of this study was to understand how participants specifically engaged with cancer prevention and screening information, searches were refined midway through the interview. The search terms *cancer prevention* and *cancer screening* were entered in all interviews approximately 30 minutes into the scrolling process to narrow the search

results. For each refined search term, the content was scrolled through up to 12 months prior and discussed as previously stated. On several occasions, when guided by the participant and the discussion at hand, additional search terms were added to find specific cancer prevention and screening information participants recalled engaging with. For example, one participant specifically recalled engaging with a post containing information about cancer and soursop (*guanábana*), a Latin American fruit commonly assumed to have curative properties. The post was elicited by searching for *cancer guanábana*. Similarly, another participant recalled a post about cancer diets and asked to search for *cancer diet*. A final search was performed using the term *cancer* and the filter *Posted by you*. This revealed any cancer information posted by the participant on their own Facebook profile.

After discussing the posts, participants were asked wrap-up questions regarding what would make cancer information more appealing on Facebook, who they considered the most influential and trustworthy sources of cancer information among their Facebook friends, and whether Facebook was a source of cancer information they trusted. Notes were taken throughout the interviews and used to inform data management and analysis.

Data Management

The data collection processes described above elicited rich data: in addition to survey responses, >20 hours of computer screen video and >30 hours of interview audio were captured (Figure 1). Survey responses were entered into a Microsoft Excel spreadsheet. Interview audio recordings were deidentified and transcribed verbatim. The process of capturing discussed posts and deidentifying data recorded on the computer screen is described in the following sections.

The first step in managing all computer screen recordings was to develop a checklist to document all the decision points for each interview video. This checklist collected the time stamps for both the audio and video versions of each interview, which allowed the research team to map interview transcripts with the discussed posts during analysis. Audio and video time stamps were collected at the beginning of the video recording and at the beginning of each post discussed. In addition to marking the time stamps for each post, the checklist was used to summarize the content of each post and to highlight relevant points discussed during the interview. These notes were incorporated as memos associated with each post during the analysis. The checklist was also used to document any search term refinements and outline preliminary codes for subsequent codebook development.

After using the checklist to document each post discussed in the interview, the post was captured through a screen grab and deidentified by cropping and/or covering any identifying images or names with white boxes and saved as a new file identified with the participant's unique ID; 2 additional files were saved in addition to the post screen grabs when applicable. First, if the post also included a video, the video was captured in its entirety in one of two ways: (1) if the video was part of a publicly available post, the lead researcher recorded the full video by searching for the post on Facebook or (2) if the video was no longer available on Facebook, the segment of the

recorded computer screen was trimmed and cropped using iMovie to ensure that the video was deidentified. Second, if a post included a link to an external website that was visited during the interview, the website was captured in one of two ways: (1) if the website link was still accessible, the lead researcher saved a web archive and PDF version of the website page or (2) if the website link was broken or no longer accessible, the recorded segment was deidentified, as described above. All deidentified files (posts, videos, web archives, surveys, and interview transcripts) were saved in a secure cloud-based file sharing and file storage service through the Johns Hopkins University and in an encrypted folder on a password-protected computer. The deidentified data were managed using the MAXQDA (Version 12; VERBI GmbH).

Data Analysis

The last step was to analyze multiple data elicited through the aforementioned methods. This was performed using traditional data analysis approaches (ie, frequencies, content analysis, and thematic analysis) that were triangulated to explain how and why engagement with cancer prevention and screening content occurred and how this engagement led to further actions. In the following sections, we summarize these analytical approaches; a detailed description of these analyses can be found in the original study [37].

First, we conducted descriptive statistics on all survey data. These findings were used to assist in contextualizing our sample. Then, a content analysis was conducted on all cancer prevention and screening information participants engaged with on their Facebook accounts. Content analysis was used to assess message patterns in a variety of formats, including those available on internet platforms [44]. A codebook was developed using the preliminary codes documented in the checklist during the data management process described in the previous section. The initial coding framework was applied to a sample of 10 cancer posts publicly available on Facebook by the lead researcher and a second bilingual study team member. Discrepancies were discussed and resolved, and a final codebook was developed [37]. Codes were developed for the following areas: post features, post source, post content, and credibility assessment. A total of 2 coders independently coded 10% of the sample. Intercoder reliability was calculated (0.89-1.0) [45], and any discrepancies were discussed until consensus was reached. The lead researcher coded the remaining posts, and code frequencies were calculated upon completion.

Finally, a thematic analysis was conducted on all the interview transcripts. This method allowed for the identification, analysis, and interpretation of patterns or themes in rich interview data sets [46,47], allowing a detailed description of how multiple themes and factors work together to explain engagement with cancer information. Transcriptions were analyzed in their original language to ensure that no meanings were lost in translation. The transcripts were preliminarily coded using emerging codes that aligned with the research questions using a constant comparison method [48]. A coding tree was created to outline the discovered themes and concepts. In addition, memos were composed with exemplary quotes for each theme; any exemplary quotes collected in Spanish were translated into

English. Memos were discussed with the study team to ensure dependability and credibility in theme development [49]. The data were placed into larger themes and factors to comprehensively explain how the phenomena occurred. Further data validation was conferred by triangulating the thematic analysis results with those of the content analysis [50] and is discussed in the original paper [37].

Ethics Approval

This study was approved by the Johns Hopkins School of Public Health institutional review board (IRB8484).

Results

Overview

Our study sample comprised 20 self-identified Latino and Latina Facebook users aged 40 to 75 years without a history of cancer (average age 54.2, SD 7.4 years) and represented 7 distinct Latin American subethnic groups from the Caribbean, Central America, and South America; 9 (45%) participants were fully bilingual, 6 (30%) preferred Spanish, and 5 (25%) preferred English. Participants were mainly female (15/20, 75%) and heavy Facebook users, with most (17/20, 85%) reporting checking their Facebook at least once a day. Facebook was most commonly used for social interaction (17/20, 85%) and information sharing (15/20, 75%). Participants had a median value of 357 (IQR 189.5-544.5) Facebook friends and followed a median of 20 (IQR 4.5-56) Facebook groups; only one of the participants followed cancer-related Facebook groups. A detailed description of the sample is available in the main study [37].

Overall, participants reported engaging with 66 posts containing cancer prevention and screening information (4.1 average posts per participant) in the previous year. Data emerging from the social media content and context elicitation method demonstrated that participants mostly engaged with cancer prevention and screening information by viewing and/or reading content (48/66, 73% posts) rather than by liking, commenting, or sharing posts (18/66, 27% posts). Furthermore, it provided rich content regarding how Latinos and Latinas engage with and act upon cancer prevention and screening information on Facebook [37]. In the following sections, we explore 2 sample cases to illustrate how a mixed methods approach provides rich insight that is otherwise missed when quantitative methods are used alone. These 2 cases were selected as they were emblematic of the broad range of information elicited from our sample that goes beyond only quantifying engagement. Participants' names have been changed to protect their identities.

Case 1: Rogelio

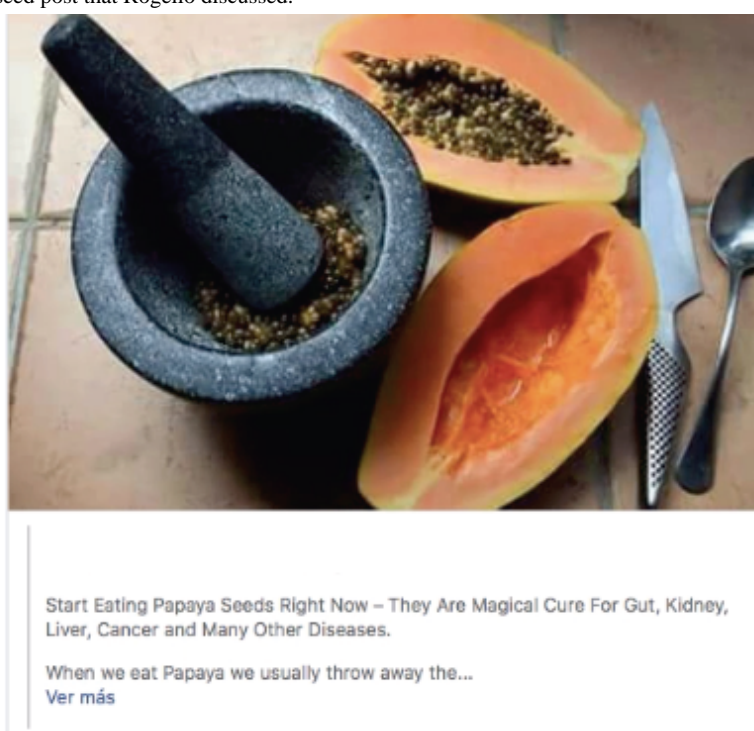
Rogelio was a bilingual Cuban male aged 61 years. He had >1800 Facebook friends and followed 131 Facebook groups, none of which were related to cancer. He considered himself a very active Facebook user, logging in multiple times a day and using the platform for social interactions, searching for and sharing information, seeing what others are doing, and maintaining his cultural identity. During the interview, 13 cancer-related posts were discussed, all of which had a video or image, for he believed that "if it doesn't enter through the eyes, it doesn't reach you." Although he engaged with all 13

posts by reading the content, he did not like, comment, or share any of these on his profile. All but 1 of these posts were shared by friends in his network; the other was shared by a Facebook group to which he belonged. A total of 6 posts were related to natural remedies or foods with curative properties against cancer, 1 was about a free skin cancer screening event, and 1 was about free colorectal and prostate cancer educational sessions for Latino men; the remaining posts were related to cancer survivorship and prayer requests.

Although Rogelio used his Facebook account frequently throughout the day, he explained that he rarely liked, commented, or shared content on his profile as he could not let others know he was on Facebook during work hours. Therefore, instead of engaging with a post through these metrics, he would send himself interesting posts through Facebook Messenger

(the platform's messaging tool). In this manner, he could read the post at a later time. He also explained how he and his wife regularly shared information related to diet and foods with preventive and/or curative properties through Facebook Messenger. Many times, after discussing content that either one engaged with on Facebook, he would decide whether they would incorporate these natural remedies into their daily lifestyle; he mentioned doing this with the 6 posts discussed during the interview. For example, he described how he and his wife started to eat papaya seeds after he read a post stating that "they are [sic] magical cure for gut, kidney, liver, cancer and many other diseases" (Figure 2). This post described how to consume papaya seeds and outlined 8 benefits, including that papaya seeds "have agents that can stop the growth of tumors and cancer cells, [and] contain isothiocyanate, which helps with breast, colon, leukemia, lung and prostate cancer."

Figure 2. Image of the papaya seed post that Rogelio discussed.



Rogelio also stated that, although Facebook was one of his main sources of information, he rarely—if ever—verified the information he engaged with on the platform. Instead, he relied on the seriousness of the people who post content on their profiles, stating that his friends from church or those aged >40 years are serious and do not share *fake news*. He also relied on his previous knowledge about a topic and believed that posts about the curative properties of foods are more credible than other topics. For Rogelio, engaging with information through a post was sufficient for him and his wife to incorporate natural remedies into their diets, regardless of whether the post cited an information source.

Finally, his cultural values and Cuban heritage came up frequently during the interview. He tended to have a fatalistic view about cancer, which emerged in multiple discussions. For example, he recalled seeing a post pertaining to 2 educational events for men about colorectal and prostate cancer. When he saw it, he immediately said he never attended such events as

speaking about these topics is like inviting the disease into your life:

It's like not wanting to speak about the topic, so it doesn't happen to me. As if talking about [colorectal or prostate cancer] puts it in my cabinet.

He believed this avoidance is a very negative Latin American custom; however, he claimed Latinos and Latinas rather *look the other way* when these topics emerge.

Case 2: Luisa

Luisa was a Puerto Rican female aged 63 years who preferred English. She had 370 Facebook friends and followed 268 Facebook groups, none of which were related to cancer. She also considered herself an avid Facebook user, logging on multiple times a day and using the platform for social interactions, searching for and sharing information, passing time, entertainment, relaxing, expressing her opinions, seeing

what others are doing, advocacy, and convenience. During the interview, 11 cancer-related posts were discussed, 5 of which contained cancer prevention and screening information she engaged with. Another 2 posts containing cancer prevention and screening information were discussed as they grabbed her attention during the interview; she had not recalled seeing them previously but stated that she would have read them if she had as they were posted by a friend who she deemed a trustworthy source of health information. The remaining discussed cancer-related posts pertained to cancer survivorship and requests for prayer for survivors of cancer. She shared only 1 post on her profile; she did not like, comment, or share any of the other posts discussed.

When discussing her Facebook use patterns, Luisa stated that she sometimes did not engage through likes, comments, or shares as she was just scrolling through her timeline and did not stop to perform these actions. However, she said this does not mean that she failed to read or watch the content. She gave an example of being at the grocery store line while scrolling through her Facebook: she might watch an interesting video but does not stop to share it with others, only sharing content when “relaxed.”

Luisa was very interested in topics pertaining to cancer prevention, particularly those related to a healthy diet. She discussed superfoods frequently and stated her preference for natural remedies over medication. For example, when discussing a video that included “10 alkaline foods that prevent and treat diabetes, gout, heart disease, and cancer,” she stated that it was the images of different superfoods that initially grabbed her attention, not the cancer prevention claims. She also mentioned that repetition surrounding the benefits of superfoods confirms the credibility of such information. She gave an example of this while discussing engagement with a post about soursop, which stated that it “has been used by many people to fight against cancer cells.” Luisa said that she was familiar with the curative properties of soursop as she had heard this often from friends and family in Puerto Rico. In fact, she had tried to incorporate it into her diet but had not been able to find it in any local supermarket.

Throughout the interview, Luisa continuously mentioned having seen a post about juicing as a way of preventing cancer. She recalled having seen the post on Facebook and copying the recipe on her phone’s notepad app. In discussing this, she also mentioned using Facebook Messenger to send herself articles. At the end of the interview, we were able to find the post by entering the search term *cancer juice*. The post claimed that the super juice recipe “is designed to help us combat breast cancer, as well as helping to starve off all potential cancer cells within the body.” It also stated that the juice cannot be blended as it is a *therapy tonic* that must be prepared using a juicer. The recipe called for broccoli, kale, cauliflower, fresh ginger root, apples, and carrots. She shared that she had since incorporated this juice into her diet, asking for it to be prepared for her when she goes to the supermarket. When asked, she said she decided to include this juice as part of her diet as she considered the friend who posted the recipe to be an extremely trustworthy source of health information. This friend came up 4 times during the interview as she often shared information about natural remedies against

many diseases on Facebook, a topic Luisa was very interested in. As Luisa considered this person a trustworthy source of information, she said she rarely further verified the content she posts and might instead just send her any questions through Facebook Messenger. She trusted that her friend had already verified the content shared, although all the websites shared by her friend lacked sources of evidence-based information. When she does decide to verify any information she finds on Facebook, she goes to Google and WebMD.

Discussion

Principal Findings

This study presented a qualitatively driven, mixed methods approach to explore how individuals engage with health information on Facebook (specifically, cancer prevention and screening information) and the impact engagement may have on subsequent behavior. In doing so, it expands upon what is known regarding cancer information engagement on social media, which predominantly stems from quantitative methodologies. The current literature operationalizes engagement with information on Facebook through likes, comments, and shares, with some studies further categorizing engagement into levels by type of engagement [16,18-20,25,31]. However, the social media content and context elicitation method adds yet another layer of nuance to public health’s current conceptualization of engagement by providing insight into the different ways people may process and act upon information, particularly individuals who would rather not like, comment, or share posts they consume. As exemplified in the aforementioned case studies, individuals may choose to read, discuss, or even change their behavior based on cancer prevention and screening information they consume without liking, commenting, and/or sharing the information. The aforementioned case studies also show that some individuals may circumvent liking, commenting, and/or sharing by using other messaging platforms to store or share information with others, such as Facebook Messenger and WhatsApp. These findings highlight the importance of exploring how platform interconnectivity affects health information engagement. As such, the presented methodology can assist in developing more comprehensive models describing engagement with health information on social media, responding to calls for a more thorough understanding of engagement on the social media landscape [15].

Consistent with previous literature [35], there are many reasons individuals do not engage with content in ways visible to others on social media. However, this decision is not indicative of a lack of engagement: both cases discussed in this study demonstrate ways in which individuals engage with and even disseminate posts while circumventing likes, comments, and shares. Discounting these aspects of engagement provides a limited explanation of the impact of health information in the social media landscape. This is of paramount importance in the current web-based environment, which is increasingly bombarded with misinformation on a broad range of topics. The social media content and context elicitation method is able to obtain a robust account of how individuals engage with health

misinformation, what grabs their attention, how they perceive it, and how they incorporate this information into their daily lives. These insights are necessary to counteract the impact misinformation may have on the uptake of cancer prevention and screening recommendations, which is a growing area of research interest [1]. Although we explore the ramifications of engagement with cancer prevention and screening misinformation in a forthcoming publication, other researchers have already adapted the social media content and context elicitation method to explore the factors related to engagement and disengagement with COVID-19 information on the web [51]. As such, the social media content and context elicitation method may be of particular interest to public health efforts developing social media campaigns targeting misinformation among populations with lower digital and/or health literacy. This method can also provide further insight into features that affect engagement and contribute to the dissemination of accurate cancer information, particularly those conveying prevention and screening recommendations. This method may also be applied to future studies regarding how to best communicate health information on these platforms, an important step toward addressing health disparities.

The process of developing this mixed methodology led to several insights. First, it is important to have a thorough understanding of the social media platform to be explored and its features to maximize how data can be accessed and used for research. In this study, understanding the features that Facebook provides when searching for content on the platform allowed the development of a detailed process to access content alongside participants that may otherwise not be accessible. It also allowed researchers to chronologically discuss content in person with participants, which overrides any algorithms that may affect the visibility of content, while also providing a glimpse to the overall cancer information landscape participants encounter on Facebook. This content not only included cancer prevention and screening information but also information about cancer survivorship, treatment, research, and other cancer topics. In fact, posts with cancer information unrelated to prevention and screening were more common than posts about cancer prevention and screening. Another important observation is that research teams must adapt to the quickly changing nature of social media platforms when embarking on such research efforts. For example, midway during data collection, it was observed that Facebook added a new filter option to their search, which enables users to look only at *Posts you've seen*. Although details on how Facebook determines which posts a person has seen are not readily available, including this filter in future research using the methods described in this paper would reduce potential participant recall bias [52].

There are also important ethical considerations researchers must take into account when developing new methodologies to explore content in an increasingly unreliable information landscape on social media. One of these considerations entails privacy concerns. This study took place several months after Facebook's Cambridge Analytica scandal, where the information of 50 million American Facebook users was used to identify voters' personalities and influence voting behaviors in the 2016 election [53]. In an additional measure of clarity, the study team

developed an additional information sheet for participants that outlined privacy expectations, what data would be captured, and what would and would not be done with captured data once deidentified. It also included images that provided an example of how the discussed posts would be deidentified before analysis. This sheet was discussed in person during the informed consent process and served as a useful resource to ensure participants fully understood the study methods and measures taken to protect the privacy of secondary data. Thus, it is important to be up to date on current events pertaining to social media platforms and issues concerning privacy and other policies that may increase perceptions of mistrust among the general public. It is also important to ensure that potential participants are extremely clear in their understanding of data safeguards in studies that use the aforementioned methods or any other mixed methodologies that capture information from a participant's social media account or accounts.

This study has several limitations. First, on a practical level, the method described is labor intensive and requires a detailed data collection and management protocol, increasing the resources needed to conduct similar research on a larger scale. This approach may also not be appropriate for more sensitive health topics or individuals who may find these in-depth methods too strenuous. Second, although participants accessed their Facebook accounts on a study laptop, 60% (12/20) of participants reported only accessing their accounts on their cell phones. The visual layout of Facebook's website version is different from that of its mobile app. This difference in visualization may have affected the ability of some participants to fully recall some posts they previously engaged with as they looked different on the computer screen. Future studies conducting this type of methodology may want to explore using a mobile device to collect data. They may also incorporate the aforementioned new *Posts you've seen* filter to minimize recall bias more generally, as self-reported recall may capture only content that people more deeply engaged with rather than all content to which they were exposed and maybe glanced over. Finally, only posts that included the search terms in the text emerged in the search during the data collection process, inevitably excluding posts that did not contain some kind of text feature (eg, posts with only a picture or a direct link to a video). It also excluded posts that discussed cancer-related topics but did not, at minimum, include the word *cancer*, whereas it included posts unrelated to the disease (eg, astrology-related posts or those equating current events in Latin American politics to cancer). Future studies should ensure they possess a comprehensive list of search terms encompassing multiple areas of the study topic while understanding that an increase in search terms adds time to the interview.

Conclusions

The social media content and context elicitation method shows potential for a deeper contextualization of engagement with health information on social media. Conducting interviews to complement the quantitative content analysis of elicited posts allows a deeper understanding of the reasons and ways engagement with health information on social media occurs, which cannot be done by observing web-based content alone [54] or by asking questions that require recall about a topic that

may not be salient to most (ie, cancer prevention and screening information engagement). This mixed methodology also allows a discussion of how message engagement may be a result of offline interactions and relationships and how these affect assessments of message credibility and accuracy. Our findings provide insight into the preferred source and content characteristics of information on social media that triggers engagement and subsequent action among specific groups and vulnerable populations, laying foundational work for the

development of future measures and empirical research exploring innovative and participatory health communication on social media platforms. Future steps for the research described in this paper include data integration and the development of a final conceptual model to help visualize the process of engagement with cancer prevention and screening information on Facebook among Latinos and Latinas in the United States.

Acknowledgments

This project was supported by Johns Hopkins Bloomberg School of Public Health Center for Qualitative Studies in Health and Medicine dissertation enhancement award and the National Cancer Institute Center to Reduce Cancer Health Disparities award #P30CA177558-05S3. The first author was previously supported by the National Cancer Institute of the National Institutes of Health (awards #T32CA009314-36, 2015-2018, and #F31CA224615, 2018-2020). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute.

The authors would like to acknowledge the assistance of Dr Javier de la Maza with content analysis coding and the transcription services of Dr Carlos Rodríguez Díaz Sex+ TEAM at the Center for Evaluation and Sociomedical Research at the University of Puerto Rico, School of Public Health–Medical Sciences Campus.

Conflicts of Interest

None declared.

References

1. Chou WS, Oh A, Klein WM. Addressing health-related misinformation on social media. *J Am Med Assoc* 2018 Dec 18;320(23):2417-2418. [doi: [10.1001/jama.2018.16865](https://doi.org/10.1001/jama.2018.16865)] [Medline: [30428002](https://pubmed.ncbi.nlm.nih.gov/30428002/)]
2. HCIRB Research Priority: misinformation in the age of social media. National Cancer Institute. 2018. URL: <https://cancercontrol.cancer.gov/brp/hcirb/misinformation.html> [accessed 2019-11-15]
3. Sell T, Hosangadi D, Smith E, Trotochaud M, Vasudevan P, Gronvall G, et al. National priorities to combat misinformation and disinformation for COVID-19 and future public health threats: a call for a national strategy. Johns Hopkins Center for Health Security, Baltimore, MD. 2021. URL: <https://www.centerforhealthsecurity.org/our-work/publications/national-priorities-to-combat-misinformation-and-disinformation-for-covid-19> [accessed 2021-10-01]
4. Chou WS, Prestin A, Lyons C, Wen K. Web 2.0 for health promotion: reviewing the current evidence. *Am J Public Health* 2013 Jan;103(1):9-18. [doi: [10.2105/AJPH.2012.301071](https://doi.org/10.2105/AJPH.2012.301071)] [Medline: [23153164](https://pubmed.ncbi.nlm.nih.gov/23153164/)]
5. Heldman AB, Schindelar J, Weaver JB. Social media engagement and public health communication: implications for public health organizations being truly “Social”. *Public Health Rev* 2013 Jun 3;35(1):13. [doi: [10.1007/bf03391698](https://doi.org/10.1007/bf03391698)]
6. 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](https://doi.org/10.1126/science.aap9559)] [Medline: [29590045](https://pubmed.ncbi.nlm.nih.gov/29590045/)]
7. Sommariva S, Vamos C, Mantzarlis A, Đào LU, Martinez Tyson D. Spreading the (Fake) news: exploring health messages on social media and the implications for health professionals using a case study. *Am J Health Edu* 2018 Jun 07;49(4):246-255. [doi: [10.1080/19325037.2018.1473178](https://doi.org/10.1080/19325037.2018.1473178)]
8. Törnberg P. Echo chambers and viral misinformation: modeling fake news as complex contagion. *PLoS One* 2018 Sep 20;13(9):e0203958 [FREE Full text] [doi: [10.1371/journal.pone.0203958](https://doi.org/10.1371/journal.pone.0203958)] [Medline: [30235239](https://pubmed.ncbi.nlm.nih.gov/30235239/)]
9. Waszak PM, Kasprzycka-Waszak W, Kubanek A. The spread of medical fake news in social media – The pilot quantitative study. *Health Policy Technol* 2018 Jun;7(2):115-118. [doi: [10.1016/j.hlpt.2018.03.002](https://doi.org/10.1016/j.hlpt.2018.03.002)]
10. Schroepfer M. An update on our plans to restrict data access on Facebook. Facebook Newsroom. 2018. URL: <https://newsroom.fb.com/news/2018/04/restricting-data-access/> [accessed 2019-11-15]
11. Cresswell JW, Clark VL. *Designing and Conducting Mixed Methods Research* (2nd ed.). Thousand Oaks, CA: Sage Publications; 2011.
12. Hughes C. It's time to break up Facebook. *New York Times*. 2019. URL: <https://www.nytimes.com/2019/05/09/opinion/sunday/chris-hughes-facebook-zuckerberg.html?searchResultPosition=13> [accessed 2019-06-15]
13. Perrin A, Anderson M. Share of U.S. adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. 2019. URL: <https://www.pewresearch.org/fact-tank/2019/04/10/share-of-u-s-adults-using-social-media-including-facebook-is-mostly-unchanged-since-2018/> [accessed 2019-06-15]
14. Alhabash S, Ma M. A tale of four platforms: motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students? *Soc Media Soc* 2017 Feb 01;3(1):205630511769154. [doi: [10.1177/2056305117691544](https://doi.org/10.1177/2056305117691544)]

15. Lim MS, Wright CJ, Carrotte ER, Pedrana AE. Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites. *Health Promot J Austr* 2016 Feb 06;27(3):187-197. [doi: [10.1071/HE16057](https://doi.org/10.1071/HE16057)] [Medline: [27719734](https://pubmed.ncbi.nlm.nih.gov/27719734/)]
16. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. *Health Promot Pract* 2013 Mar;14(2):157-162. [doi: [10.1177/1524839912469378](https://doi.org/10.1177/1524839912469378)] [Medline: [23271716](https://pubmed.ncbi.nlm.nih.gov/23271716/)]
17. 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](https://doi.org/10.1080/10810730.2016.1271065)] [Medline: [28248623](https://pubmed.ncbi.nlm.nih.gov/28248623/)]
18. Srivastava J, Saks J, Weed AJ, Atkins A. Engaging audiences on social media: identifying relationships between message factors and user engagement on the American Cancer Society's Facebook page. *Telemat Informat* 2018 Oct;35(7):1832-1844. [doi: [10.1016/j.tele.2018.05.011](https://doi.org/10.1016/j.tele.2018.05.011)]
19. Biancovilli P, Jurberg C. How to optimize health messages about cancer on Facebook: mixed-methods study. *JMIR Cancer* 2018 Dec 18;4(2):e11073 [FREE Full text] [doi: [10.2196/11073](https://doi.org/10.2196/11073)] [Medline: [30563821](https://pubmed.ncbi.nlm.nih.gov/30563821/)]
20. Biancovilli P, Picanço L, Jurberg C. To read or not to read? Identifying communication patterns in three cancer-related Facebook pages. *Cogent Soc Sci* 2017 May 24;3(1):1331816. [doi: [10.1080/23311886.2017.1331816](https://doi.org/10.1080/23311886.2017.1331816)]
21. Strekalova YA, Krieger JL. Beyond words: amplification of cancer risk communication on social media. *J Health Commun* 2017 Oct;22(10):849-857. [doi: [10.1080/10810730.2017.1367336](https://doi.org/10.1080/10810730.2017.1367336)] [Medline: [28956723](https://pubmed.ncbi.nlm.nih.gov/28956723/)]
22. Strekalova YA, Damiani RE. Message design and audience engagement with tobacco prevention posts on social media. *J Cancer Educ* 2018 Jun;33(3):668-672. [doi: [10.1007/s13187-016-1135-x](https://doi.org/10.1007/s13187-016-1135-x)] [Medline: [27832508](https://pubmed.ncbi.nlm.nih.gov/27832508/)]
23. Theiss SK, Burke RM, Cory JL, Fairley TL. Getting beyond impressions: an evaluation of engagement with breast cancer-related Facebook content. *Mhealth* 2016 Nov;2:41 [FREE Full text] [doi: [10.21037/mhealth.2016.10.02](https://doi.org/10.21037/mhealth.2016.10.02)] [Medline: [27840816](https://pubmed.ncbi.nlm.nih.gov/27840816/)]
24. Strekalova YA, Krieger JL. A picture really is worth a thousand words: public engagement with the National Cancer Institute on social media. *J Cancer Educ* 2017 Mar 8;32(1):155-157 [FREE Full text] [doi: [10.1007/s13187-015-0901-5](https://doi.org/10.1007/s13187-015-0901-5)] [Medline: [26351003](https://pubmed.ncbi.nlm.nih.gov/26351003/)]
25. Klippert H, Schaper A. Using Facebook to communicate mammography messages to rural audiences. *Public Health Nurs* 2019 Mar;36(2):164-171. [doi: [10.1111/phn.12556](https://doi.org/10.1111/phn.12556)] [Medline: [30370687](https://pubmed.ncbi.nlm.nih.gov/30370687/)]
26. Huesch M, Chetlen A, Segel J, Schetter S. Frequencies of private mentions and sharing of mammography and breast cancer terms on Facebook: a pilot study. *J Med Internet Res* 2017 Jun 09;19(6):e201 [FREE Full text] [doi: [10.2196/jmir.7508](https://doi.org/10.2196/jmir.7508)] [Medline: [28600279](https://pubmed.ncbi.nlm.nih.gov/28600279/)]
27. Kashian N, Jacobson S. Factors of engagement and patient-reported outcomes in a Stage IV Breast Cancer Facebook Group. *Health Commun* 2020 Jan;35(1):75-82. [doi: [10.1080/10410236.2018.1536962](https://doi.org/10.1080/10410236.2018.1536962)] [Medline: [30351185](https://pubmed.ncbi.nlm.nih.gov/30351185/)]
28. Valle CG, Tate DF. Engagement of young adult cancer survivors within a Facebook-based physical activity intervention. *Transl Behav Med* 2017 Dec;7(4):667-679 [FREE Full text] [doi: [10.1007/s13142-017-0483-3](https://doi.org/10.1007/s13142-017-0483-3)] [Medline: [28374211](https://pubmed.ncbi.nlm.nih.gov/28374211/)]
29. Stapleton JL, Manne SL, Day AK, Levonyan-Radloff K, Pagoto SL. Healthy body image intervention delivered to young women via Facebook groups: formative study of engagement and acceptability. *JMIR Res Protoc* 2018 Feb 20;7(2):e54 [FREE Full text] [doi: [10.2196/resprot.9429](https://doi.org/10.2196/resprot.9429)] [Medline: [29463495](https://pubmed.ncbi.nlm.nih.gov/29463495/)]
30. Cole-Lewis H, Perotte A, Galica K, Dreyer L, Griffith C, Schwarz M, et al. Social network behavior and engagement within a smoking cessation Facebook page. *J Med Internet Res* 2016 Aug 02;18(8):e205 [FREE Full text] [doi: [10.2196/jmir.5574](https://doi.org/10.2196/jmir.5574)] [Medline: [27485315](https://pubmed.ncbi.nlm.nih.gov/27485315/)]
31. Wang X, Chen L, Shi J, Peng T. What makes cancer information viral on social media? *Comput Hum Behav* 2019 Apr;93:149-156. [doi: [10.1016/j.chb.2018.12.024](https://doi.org/10.1016/j.chb.2018.12.024)]
32. How does Facebook determine which videos are Suggested for You? Facebook Help Center. URL: <https://www.facebook.com/help/iphone-app/1202019566532169> [accessed 2019-06-15]
33. Sethuraman R, Vallmitjana J, Levin J. Using surveys to make news feed more personal. Facebook Newsroom. 2019. URL: <https://newsroom.fb.com/news/2019/05/more-personalized-experiences/> [accessed 2019-06-15]
34. Edelman N. Reviewing the definitions of "lurkers" and some implications for online research. *Cyberpsychol Behav Soc Netw* 2013 Sep;16(9):645-649. [doi: [10.1089/cyber.2012.0362](https://doi.org/10.1089/cyber.2012.0362)] [Medline: [23848960](https://pubmed.ncbi.nlm.nih.gov/23848960/)]
35. Sun N, Rau PP, Ma L. Understanding lurkers in online communities: a literature review. *Comput Hum Behav* 2014 Sep;38:110-117. [doi: [10.1016/j.chb.2014.05.022](https://doi.org/10.1016/j.chb.2014.05.022)]
36. Bleakley W. About us: learn more about CrowdTangle. CrowdTangle. URL: <https://help.crowdtangle.com/en/articles/4201940-about-us> [accessed 2021-07-15]
37. Rivera YM, Moran MB, Thrul J, Joshu C, Smith KC. When engagement leads to action: understanding the impact of cancer (mis)information among Latino/a Facebook users. *Health Commun* 2021 Jul 13:1-13. [doi: [10.1080/10410236.2021.1950442](https://doi.org/10.1080/10410236.2021.1950442)] [Medline: [34256676](https://pubmed.ncbi.nlm.nih.gov/34256676/)]
38. Xu J, Murphy SL, Kochanek KD, Bastian BA. Deaths: final data for 2013. *Natl Vital Stat Rep* 2016 Feb 16;64(2):1-119 [FREE Full text] [Medline: [26905861](https://pubmed.ncbi.nlm.nih.gov/26905861/)]

39. Cancer facts and figures for Hispanics/Latinos 2018-2020. American Cancer Society. 2018. URL: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/cancer-facts-and-figures-for-hispanics-and-latinos/cancer-facts-and-figures-for-hispanics-and-latinos-2018-2020.pdf> [accessed 2019-12-20]
40. Boyce C, Neale P. Conduction in-depth interviews: a guide for designing and conducting in-depth interviews for evaluation input. Pathfinder International Tool Series: Monitoring and Evaluation 2. 2006. URL: <http://compositionawebb.pbworks.com/f/> [accessed 2019-01-01]
41. Rubin A. Uses and gratifications: an evolving perspective of media effects. In: Nabi RL, Oliver MB, editors. The SAGE Handbook of Media Processes and Effects. Thousand Oaks, CA: SAGE Publications; 2010.
42. Johnson J, Andrews JE, Allard S. A model for understanding and affecting cancer genetics information seeking. Library Inform Sci Res 2001 Dec;23(4):335-349. [doi: [10.1016/s0740-8188\(01\)00094-9](https://doi.org/10.1016/s0740-8188(01)00094-9)]
43. Johnson JD. Cancer-related Information Seeking (Health Communication). Cresskill, NJ: Hampton Press; 1997.
44. Manganello J, Fishbein M. Using theory to inform content analysis. In: Jordan AB, Kunkel D, Manganello J, Fishbein M, editors. Media Messages and Public Health: A Decisions Approach to Content Analysis. New York: Routledge; 2009:3-14.
45. Neuendorf K. Reliability for content analysis. In: Jordan AB, Kunkel D, Manganello J, Fishbein M, editors. Media Messages and Public Health: A Decisions Approach to Content Analysis. New York: Routledge; 2009:67-87.
46. Boyatzis R. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: SAGE Publications; 1998.
47. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: [10.1191/1478088706qp063oa](https://doi.org/10.1191/1478088706qp063oa)]
48. Dye J, Schatz I, Rosenberg B, Coleman S. Constant comparison method: a kaleidoscope of data. Qual Rep 2015 Feb 2;4(1):1-10. [doi: [10.46743/2160-3715/2000.2090](https://doi.org/10.46743/2160-3715/2000.2090)]
49. Guba E, Lincoln Y. Fourth Generation Evaluation. Newbury Park, CA: SAGE Publications; 1989.
50. Morse JM. Critical analysis of strategies for determining rigor in qualitative inquiry. Qual Health Res 2015 Sep;25(9):1212-1222. [doi: [10.1177/1049732315588501](https://doi.org/10.1177/1049732315588501)] [Medline: [26184336](https://pubmed.ncbi.nlm.nih.gov/26184336/)]
51. Houlden S, Hodson J, Veletsianos G, Thompson CP, Reid D. Inoculating an infodemic: an ecological approach to understanding engagement with COVID-19 online information. Am Behav Scientist 2021 Oct 18;65(14):1990-2013. [doi: [10.1177/00027642211050903](https://doi.org/10.1177/00027642211050903)]
52. Niederdeppe J. Assessing the validity of confirmed AD recall measures for public health communication campaign evaluation. J Health Commun 2005;10(7):635-650. [doi: [10.1080/10810730500267662](https://doi.org/10.1080/10810730500267662)] [Medline: [16278200](https://pubmed.ncbi.nlm.nih.gov/16278200/)]
53. Granville K. Facebook and Cambridge Analytica: what you need to know as fallout widens. New York Times. 2018. URL: <https://www.nytimes.com/2018/03/19/technology/facebook-cambridge-analytica-explained.html> [accessed 2019-10-15]
54. Seale C, Charteris-Black J, MacFarlane A, McPherson A. Interviews and internet forums: a comparison of two sources of qualitative data. Qual Health Res 2010 May;20(5):595-606. [doi: [10.1177/1049732309354094](https://doi.org/10.1177/1049732309354094)] [Medline: [20008955](https://pubmed.ncbi.nlm.nih.gov/20008955/)]

Edited by G Eysenbach; submitted 25.10.20; peer-reviewed by J Sanchez, E Walsh-Buhi, X Wang, J Segel; comments to author 08.12.20; revised version received 07.08.21; accepted 27.12.21; published 04.03.22.

Please cite as:

Rivera YM, Moran MB, Thrul J, Joshu C, Smith KC

Contextualizing Engagement With Health Information on Facebook: Using the Social Media Content and Context Elicitation Method
J Med Internet Res 2022;24(3):e25243

URL: <https://www.jmir.org/2022/3/e25243>

doi: [10.2196/25243](https://doi.org/10.2196/25243)

PMID: [35254266](https://pubmed.ncbi.nlm.nih.gov/35254266/)

©Yonaira M Rivera, Meghan B Moran, Johannes Thrul, Corinne Joshu, Katherine C Smith. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 04.03.2022. 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.

Letter to the Editor

Toward a Better Understanding of Quality Social Connections. Comment on “Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis”

Huachu Deng¹, MD; Xingan Qin¹, MD

Department of Gastrointestinal and Gland Surgery, First Affiliated Hospital of Guangxi Medical University, Nanning, China

Corresponding Author:

Xingan Qin, MD

Department of Gastrointestinal and Gland Surgery

First Affiliated Hospital of Guangxi Medical University

6 Shuangyong Road

Qingxiu District

Nanning, 530000

China

Phone: 86 15277114948

Email: drqinxingan@126.com

Related Articles:

Comment on: <https://www.jmir.org/2021/12/e26584>

Comment in: <https://www.jmir.org/2022/3/e37440>

Abstract

(J Med Internet Res 2022;24(3):e36739) doi:10.2196/36739

KEYWORDS

mental health; digital interventions; young people; quality social connection; depression; anxiety; systematic review; meta-analysis; patient and public involvement; mobile phone

We read the paper by Dewa et al [1] with interest. The authors performed a meta-analysis to conceptualize, appraise, and synthesize evidence on quality social connection (QSC) within digital interventions (D-QSC) and the impact on depression and anxiety outcomes for young people aged 14-24 years. They demonstrated that “D-QSC is an important and underconsidered component for youth depression and anxiety outcomes. Researchers and developers should consider targeting improved QSC between clinicians and young people within digital interventions for depression. Future research should build on our framework to further examine relationships among individual attributes of QSC, various digital interventions, and different populations.” After carefully reading, I wish to put forth the following suggestions.

Two studies [2,3] in Table 2 (“Data extraction and quality assessment of included studies”) were created by the same author team (Radovic et al) with similar characteristics (year, country, study design, setting and participants, digital intervention, and outcomes and measures). This duplicate inclusion of data would affect the credibility of the final results of the meta-analysis. I recommend that the authors exclude duplicate works in the meta-analysis and that a correction and reliable checking of the data insertion is required.

Due to these mistakes, I suggest that the authors further refine the inclusion criteria for the included studies to avoid duplication of inclusion.

Acknowledgments

We are supported by the Department of Gastrointestinal and Gland Surgery at the First Affiliated Hospital of Guangxi Medical University.

Conflicts of Interest

None declared.

References

1. Dewa LH, Lawrance E, Roberts L, Brooks-Hall E, Ashrafian H, Fontana G, et al. Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis. *J Med Internet Res* 2021 Dec 17;23(12):e26584 [FREE Full text] [doi: [10.2196/26584](https://doi.org/10.2196/26584)] [Medline: [34927592](https://pubmed.ncbi.nlm.nih.gov/34927592/)]
2. Radovic A, DeMand AL, Gmelin T, Stein BD, Miller E. SOVA: Design of a stakeholder informed social media website for depressed adolescents and their parents. *J Technol Hum Serv* 2018;35(3):169-182 [FREE Full text] [doi: [10.1080/15228835.2017.1347552](https://doi.org/10.1080/15228835.2017.1347552)] [Medline: [29743822](https://pubmed.ncbi.nlm.nih.gov/29743822/)]
3. Radovic A, Gmelin T, Stein BD, Miller E. Depressed adolescents' positive and negative use of social media. *J Adolesc* 2017 Feb;55:5-15 [FREE Full text] [doi: [10.1016/j.adolescence.2016.12.002](https://doi.org/10.1016/j.adolescence.2016.12.002)] [Medline: [27997851](https://pubmed.ncbi.nlm.nih.gov/27997851/)]

Abbreviations

D-QSC: quality social connection within digital interventions

QSC: quality social connection

Edited by T Leung; submitted 23.01.22; this is a non-peer-reviewed article; accepted 24.01.22; published 11.03.22.

Please cite as:

Deng H, Qin X

Toward a Better Understanding of Quality Social Connections. Comment on "Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis"

J Med Internet Res 2022;24(3):e36739

URL: <https://www.jmir.org/2022/3/e36739>

doi: [10.2196/36739](https://doi.org/10.2196/36739)

PMID: [35275073](https://pubmed.ncbi.nlm.nih.gov/35275073/)

©Huachu Deng, Xingan Qin. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 11.03.2022. 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.

Letter to the Editor

The Unclear Role of the Physician on Social Media During the COVID-19 Pandemic. Comment on “Emergency Physician Twitter Use in the COVID-19 Pandemic as a Potential Predictor of Impending Surge: Retrospective Observational Study”

Spencer MacLeod¹, BS; Nikhi Paul Singh¹, BS; Carter Joseph Boyd², MD, MBA

¹School of Medicine, The University of Alabama at Birmingham, Birmingham, AL, United States

²Hansjörg Wyss Department of Plastic Surgery, NYU Langone, New York City, NY, United States

Corresponding Author:

Nikhi Paul Singh, BS

School of Medicine

The University of Alabama at Birmingham

1670 University Blvd

Birmingham, AL, 35233

United States

Phone: 1 205 563 3208

Email: nsingh@uab.edu

Related Article:

Comment on: <https://www.jmir.org/2021/7/e28615>

Abstract

(*J Med Internet Res* 2022;24(3):e34870) doi:[10.2196/34870](https://doi.org/10.2196/34870)

KEYWORDS

COVID-19 pandemic; emergency medicine; disaster medicine; crisis standards of care; latent Dirichlet allocation; topic modeling; Twitter; sentiment analysis; surge capacity; physician wellness; social media; internet; infodemiology; COVID-19

We congratulate Margus and colleagues on their interesting study documenting emergency medicine physicians' use of Twitter preceding surges in COVID-19 cases [1]. The correlations discovered between tweet count and hospital case numbers represent a unique instrument to assess epidemiologic trends related to the COVID-19 pandemic [1]. An additional subanalysis by geographic region may provide enhanced insight into the efficacy of social media utilization as a predictive tool for emergency medical resource allocation. We commend the authors for their extensive search criteria employed to accurately identify US emergency medical physicians; however, this study's complex methodology highlights the lack of an official role and verification process for physicians on social media.

On Twitter, there is no specific criteria listed for verification of medical professionals. Instead, verification is based on notability criteria associated with representing a notable individual or brand [2]. These criteria exclude any physicians without a sufficiently large following from being verified on Twitter,

making it more difficult for Twitter users to identify legitimate medical professionals lacking a large following. As health professionals and researchers are further encouraged to utilize social media for professional purposes, additional verification criteria specific to medical professionals may prove beneficial in the future [3].

In 2010, the American Medical Association issued guidelines regarding the ethical use policy to aid physicians in navigating social media [4]. However, the role of physicians on these online platforms has not been uniformly described. Physicians may interact with one another, and this is evident with Med Twitter, an open-source, decentralized forum for information sharing, medical education, and professional networking, as well as the increasing use of social media in residency recruitment [5]. The observed correlation by the authors of this study may be indicative of such communication by emergency medicine physicians. Continued use of such platforms across medicine may reveal additional relationships such as this predictive model.

Perhaps a potential role may be to extend physicians' professional impact in education and patient advocacy.

As time moves forward, the use of social media in medicine will continue to expand beyond prediction as will its potential pitfalls. Margus and colleagues' article [1] brings up a positive

prospect of social media in medicine, and it is important that physicians understand the current limitations of these innovative platforms. We believe that physicians need medicine-based verification for social media in addition to clearly established roles for physicians from national governing bodies.

Conflicts of Interest

None declared.

Editorial Notice

The corresponding author of "Emergency Physician Twitter Use in the COVID-19 Pandemic as a Potential Predictor of Impending Surge: Retrospective Observational Study" declined to respond to this letter.

References

1. Margus C, Brown N, Hertelendy AJ, Safferman MR, Hart A, Ciottone GR. Emergency Physician Twitter Use in the COVID-19 Pandemic as a Potential Predictor of Impending Surge: Retrospective Observational Study. *J Med Internet Res* 2021 Jul 14;23(7):e28615 [FREE Full text] [doi: [10.2196/28615](https://doi.org/10.2196/28615)] [Medline: [34081612](https://pubmed.ncbi.nlm.nih.gov/34081612/)]
2. Twitter verification requirements - how to get the blue check. Twitter. URL: <https://help.twitter.com/en/managing-your-account/about-twitter-verified-accounts> [accessed 2021-10-10]
3. 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](https://doi.org/10.2196/13687)] [Medline: [31719028](https://pubmed.ncbi.nlm.nih.gov/31719028/)]
4. Professionalism in the use of social media. American Medical Association. URL: <https://www.ama-assn.org/delivering-care/ethics/professionalism-use-social-media> [accessed 2021-10-20]
5. Fieger E, Fadel K. #MedTwitter: An Open Source Medical Community. *MedEdPublish* 2021;10(1). [doi: [10.15694/mep.2021.000019.1](https://doi.org/10.15694/mep.2021.000019.1)]

Edited by T Leung; submitted 11.11.21; this is a non-peer-reviewed article; accepted 03.02.22; published 02.03.22.

Please cite as:

MacLeod S, Singh NP, Boyd CJ

The Unclear Role of the Physician on Social Media During the COVID-19 Pandemic. Comment on "Emergency Physician Twitter Use in the COVID-19 Pandemic as a Potential Predictor of Impending Surge: Retrospective Observational Study"

J Med Internet Res 2022;24(3):e34870

URL: <https://www.jmir.org/2022/3/e34870>

doi: [10.2196/34870](https://doi.org/10.2196/34870)

PMID: [35120018](https://pubmed.ncbi.nlm.nih.gov/35120018/)

©Spencer MacLeod, Nikhi Paul Singh, Carter Joseph Boyd. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 02.03.2022. 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.

Letter to the Editor

Interpretation Bias Toward the Positive Impacts of Digital Interventions in Health Care. Comment on “Value of the Electronic Medical Record for Hospital Care: Update From the Literature”

Erfan Shakibaei Bonakdeh^{1,2}, BSc, MS

¹Department of Management, Monash Business School, Monash University, Melbourne, VIC, Australia

²Pharmacy Department, Alfred Health, Melbourne, VIC, Australia

Corresponding Author:

Erfan Shakibaei Bonakdeh, BSc, MS

Department of Management

Monash Business School

Monash University

900 Dandenong Rd

Caulfield East

Melbourne, VIC, 3145

Australia

Phone: 61 1399032000

Email: erfan.shakibaeibonakdeh@monash.edu

Related Articles:

Comment on: <https://www.jmir.org/2021/12/e26323/>

Comment in: <https://www.jmir.org/2022/3/e37419>

Abstract

(*J Med Internet Res* 2022;24(3):e37208) doi:[10.2196/37208](https://doi.org/10.2196/37208)

KEYWORDS

cost analysis; costs and cost analyses; economic advantage; electronic medical records; electronic records; health care; hospitals; computerized medical records system; quality of health care; secondary data

The review paper by Uslu and Stausberg [1] certainly sheds some light on the positive impacts of electronic medical records (EMRs) in the hospital context. The emergence of such papers, as fruitful as they could be for minimizing skepticism among hospital executives and managers and compelling them to embark on the digitization journey, could introduce hasty and immature uptake of the technology, if biased. As a researcher heavily focused on health care digitization, I can debate that the results presented in the format of a review in the said paper are not impartial.

As highlighted by the literature, the improvement of processes caused by EMRs, as valuable as it is, may not contribute to patient outcome criteria (ie, mortality rate), and this has been one of the lengthiest debates in the field of digital health [2]. Thus, the authors' statement “the review also showed improvements in quality of care by all respective studies” struck

me as a great surprise. Further examination of the paper has brought to light that this statement was overpowered by some flaws in the study.

The bias toward declaring positive results from studies that either lacked or presented statistically insignificant positive outcomes (as noted in the original paper) is the major downside of this review. For example, Uslu and Stausberg [1] noted a positive association between EMR adoption and efficiency in the study by Adler-Milstein et al [3]. Surprisingly, in the original study, the authors clearly declared no significant association with regard to efficiency.

Additionally, Uslu and Stausberg [1] did not draw a clear line between the types of quality criteria (safety, timeliness, process, and patient outcomes), which not only is confusing to readers but also does not demonstrate the magnitude of improvement

in each dimension. As such, a criterion such as mortality rate that does not normally show a significant improvement would be overpowered by process outcomes, which often behave reversely.

The aim of the study to “summarize empirical studies about the value of electronic medical records (EMRs) for hospital care” does not justify the inclusion of a few studies [1]. For example, “Higher rates of adoption of key EHR functions among high-quality hospitals” was reported as the result of Elnahal et al [4]. The aim of the said study can show the association between the presence of high quality in targeted hospitals and the presence of IT (information technology). Thus, it is not clear if high quality was derived by the EMR or whether high-quality

hospitals adopted EMRs to maintain their status as a high-quality hospital. As the authors noted, “high quality and EHR adoption may be linked”; however, this is no strong evidence on which review studies can rely. On the other hand, the exclusion of computerized physician order entry (CPOE) was not explained by the authors as many EMRs already incorporate CPOE functions. By contrast, some studies included in the review, for example, Elnahal et al [4], mentioned the existence of CPOE in most high-quality hospitals in their sample.

To conclude, since the outcomes of secondary studies are often consulted by managers and politicians in the health care sector, researchers must be vigilant of the extensive impacts of research bias on fundamental decisions that they may cause.

Conflicts of Interest

None declared.

References

1. Uslu A, Stausberg J. Value of the Electronic Medical Record for Hospital Care: Update From the Literature. *J Med Internet Res* 2021 Dec 23;23(12):e26323 [FREE Full text] [doi: [10.2196/26323](https://doi.org/10.2196/26323)] [Medline: [34941544](https://pubmed.ncbi.nlm.nih.gov/34941544/)]
2. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. *NPJ Digit Med* 2020 Feb;3:17 [FREE Full text] [doi: [10.1038/s41746-020-0221-y](https://doi.org/10.1038/s41746-020-0221-y)] [Medline: [32047862](https://pubmed.ncbi.nlm.nih.gov/32047862/)]
3. Adler-Milstein J, Everson J, Lee SYD. EHR Adoption and Hospital Performance: Time-Related Effects. *Health Serv Res* 2015 Dec;50(6):1751-1771 [FREE Full text] [doi: [10.1111/1475-6773.12406](https://doi.org/10.1111/1475-6773.12406)] [Medline: [26473506](https://pubmed.ncbi.nlm.nih.gov/26473506/)]
4. Elnahal SM, Joynt KE, Bristol SJ, Jha AK. Electronic health record functions differ between best and worst hospitals. *Am J Manag Care* 2011 Apr;17(4):e121-e147 [FREE Full text] [Medline: [21774097](https://pubmed.ncbi.nlm.nih.gov/21774097/)]

Abbreviations

CPOE: computerized physician order entry
EMR: electronic medical record
IT: information technology

Edited by T Leung; submitted 10.02.22; this is a non-peer-reviewed article; accepted 21.02.22; published 04.03.22.

Please cite as:

Shakibaei Bonakdeh E

Interpretation Bias Toward the Positive Impacts of Digital Interventions in Health Care. Comment on “Value of the Electronic Medical Record for Hospital Care: Update From the Literature”

J Med Internet Res 2022;24(3):e37208

URL: <https://www.jmir.org/2022/3/e37208>

doi: [10.2196/37208](https://doi.org/10.2196/37208)

PMID: [35254276](https://pubmed.ncbi.nlm.nih.gov/35254276/)

©Erfan Shakibaei Bonakdeh. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 04.03.2022. 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.

Letter to the Editor

Authors' Reply to: Interpretation Bias Toward the Positive Impacts of Digital Interventions in Health Care. Comment on "Value of the Electronic Medical Record for Hospital Care: Update From the Literature"

Jürgen Stausberg¹, MD, Prof Dr; Aykut Uslu², PhD

¹Institute for Medical Informatics, Biometry and Epidemiology, University Hospital Essen, University Duisburg-Essen, Essen, Germany

²USLU Medizininformatik, Düsseldorf, Germany

Corresponding Author:

Jürgen Stausberg, MD, Prof Dr

Institute for Medical Informatics, Biometry and Epidemiology

University Hospital Essen

University Duisburg-Essen

Hufelandstrasse 55

Essen, 45122

Germany

Phone: 49 201 723 77201

Fax: 49 201 723 77333

Email: stausberg@ekmed.de

Related Articles:

Comment on: <https://www.jmir.org/2021/12/e26323/>

Comment on: <https://www.jmir.org/2022/3/e37208>

Abstract

(*J Med Internet Res* 2022;24(3):e37419) doi:[10.2196/37419](https://doi.org/10.2196/37419)

KEYWORDS

cost analysis; costs and cost analyses; economic advantage; electronic medical records; electronic records; health care; hospitals; computerized medical records systems; quality of health care; secondary data

We thank Shakibaei Bonakdeh [1] for the critical comment on the positive impact of digital interventions. It would be the ultimate success of medical informatics research to be recognized by hospital executives and managers. However, we are not convinced that the previous national initiatives for the implementation of electronic records were evidence-based, neither in England, Germany, nor the United States. Following the scene of electronic records for 30 years, recognizing the literature of 50 years, and having been responsible for the selection and management of electronic records in hospitals, our view was truthfully impartial. We were willing to accept scientific evidence independently of pre-existing opinions even if the clear result was a surprise [2].

First, we want to clarify that we did not evaluate digital interventions in general, as indicated by the letter's headline. In our study, we sought to focus on electronic medical records (EMRs) as specific types of electronic records. However, we were confronted with several challenges concerning the definition and specification of the technology, as partly addressed in the letter. For example, excluding studies focusing on computerized physician order entry (CPOE) on the one hand did not mean excluding studies on EMRs that offer CPOE support on the other hand. It would be a step forward to have a standard not only for the reporting of results such as PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) but also for the labeling of digital interventions.

Concerning the results of the included studies, we aimed to draw the utmost benefits from the details. This could mean overriding the studies' conclusions, as stated in our limitations. In the case of Adler-Milstein et al [3], we probably overweighted the improved efficiency in the post meaningful use period (2010/2011). We apologize if we did not meet the common appraisal in all cases. However, the results of Adler-Milstein et al [3] fully support the conclusions from the whole set of included studies with a clear positive effect on the quality of care and an ambiguous economical implication. The appropriate evaluation criteria of EMRs should be put up for discussion. It may be unreasonable to expect a reduction of mortality from its implementation. We would be honored if our series of reviews contribute to realistic expectations toward the effects of EMRs on different types of quality criteria.

Second, the other points of criticism were related to the context of EMR implementation. Studies about the effects of EMRs are faced with a double complexity [4]. The implementation of an EMR will involve the whole organization of hospital care, will alter health care processes, and will include a wide range of functionalities (EMR as a complex intervention). The setting of the implementation of an EMR is complex too, with different professions, different specialties, different levels of care, etc (complexity of the context). For example, the readiness for change and managing change might be more important for the

success of an EMR implementation than specific technological issues [5]. It could be argued that well-prepared organizations will benefit more from an EMR than less prepared organizations. In extreme cases, less prepared organizations will get worse with the same EMR solution that helped well-prepared organizations to further improve patient outcomes. We agree with Shakibaei Bonakdeh [1] that there could be a coexistence effect in studies using secondary data. Hospitals attempting to improve care by implementing an EMR should be advised to analyze thoroughly their current state, to eliminate reasons for inappropriate care in advance, and to be well prepared for the technology.

Medical informatics science is confronted with the digitization of health care independently from its input and participation. Due to the high penetration of electronic records, interventional studies will no longer be possible for this technology in developed countries. Nevertheless, at times, society will seek help from medical informatics science, especially if political expectations fail. Frequently, low-hanging fruits determined the reaction of researchers in the past. It would be a major step forward if medical informatics science is willing to act as a collective community grounded on scientific evidence. With this regard, we express our appreciation for Shakibaei Bonakdeh's [1] feedback.

Conflicts of Interest

None declared.

References

1. Shakibaei Bonakdeh E. Interpretation Bias Toward the Positive Impacts of Digital Interventions in Health Care. Comment on "Value of the Electronic Medical Record for Hospital Care: Update From the Literature". J Med Internet Res 2022 Mar 3;24(3):e37208. [doi: [10.2196/37208](https://doi.org/10.2196/37208)]
2. Uslu A, Stausberg J. Value of the Electronic Medical Record for Hospital Care: Update From the Literature. J Med Internet Res 2021 Dec 23;23(12):e26323 [FREE Full text] [doi: [10.2196/26323](https://doi.org/10.2196/26323)] [Medline: [34941544](https://pubmed.ncbi.nlm.nih.gov/34941544/)]
3. Adler-Milstein J, Everson J, Lee SYD. EHR Adoption and Hospital Performance: Time-Related Effects. Health Serv Res 2015 Dec;50(6):1751-1771 [FREE Full text] [doi: [10.1111/1475-6773.12406](https://doi.org/10.1111/1475-6773.12406)] [Medline: [26473506](https://pubmed.ncbi.nlm.nih.gov/26473506/)]
4. Schrappe M, Pfaff H. Health Services Research Faces New Challenges: Consequences for Definition and Concept [in German]. Gesundheitswesen 2016 Nov;78(11):689-694. [doi: [10.1055/s-0042-116230](https://doi.org/10.1055/s-0042-116230)] [Medline: [27756086](https://pubmed.ncbi.nlm.nih.gov/27756086/)]
5. Lorenzi NM, Riley RT. Managing Change: an Overview. J Am Med Inform Assoc 2000 Mar;7(2):116-124 [FREE Full text] [doi: [10.1136/jamia.2000.0070116](https://doi.org/10.1136/jamia.2000.0070116)] [Medline: [10730594](https://pubmed.ncbi.nlm.nih.gov/10730594/)]

Abbreviations

CPOE: computerized physician order entry

EMR: electronic medical record

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

Edited by T Leung; submitted 20.02.22; this is a non-peer-reviewed article; accepted 21.02.22; published 04.03.22.

Please cite as:

Stausberg J, Uslu A

Authors' Reply to: Interpretation Bias Toward the Positive Impacts of Digital Interventions in Health Care. Comment on "Value of the Electronic Medical Record for Hospital Care: Update From the Literature"

J Med Internet Res 2022;24(3):e37419

URL: <https://www.jmir.org/2022/3/e37419>

doi: [10.2196/37419](https://doi.org/10.2196/37419)

PMID: [35254272](https://pubmed.ncbi.nlm.nih.gov/35254272/)

©Jürgen Stausberg, Aykut Uslu. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 04.03.2022. 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.

Letter to the Editor

Authors' Reply to: Toward a Better Understanding of Quality Social Connections. Comment on "Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis"

Lindsay H Dewa^{1,2}, BA, MSc, PhD; Lily Roberts¹, BSc, PGCert; Emma Lawrance^{1,3}, BSc, Grad Dip, MSc, DPhil; Hutan Ashrafian¹, BSc, MBBS, MRCS, PhD

¹Institute of Global Health Innovation, Imperial College London, London, United Kingdom

²School of Public Health, Imperial College London, London, United Kingdom

³Mental Health Innovations, London, United Kingdom

Corresponding Author:

Lindsay H Dewa, BA, MSc, PhD
Institute of Global Health Innovation
Imperial College London
Reynolds Building, 3rd Floor
London, W6 8RP
United Kingdom
Phone: 44 02075940815
Email: l.dewa@imperial.ac.uk

Related Articles:

Comment on: <https://www.jmir.org/2021/12/e26584>

Comment on: <https://www.jmir.org/2022/2/e36739>

Abstract

(*J Med Internet Res* 2022;24(3):e37440) doi:[10.2196/37440](https://doi.org/10.2196/37440)

KEYWORDS

mental health; digital interventions; young people; quality social connection; depression; anxiety; systematic review; meta-analysis; patient and public involvement; mobile phone

We welcome feedback from the authors Deng and Qin through their comment [1] on our paper "Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis" [2]. They suggest that two included studies [3,4] should not be included in our review as individual studies as data extraction and quality assessment of both studies is the same, and that this impacts the quality of the meta-analysis. We have the following response. First, our review is scoping in nature, therefore we included and extracted any study that matched our inclusion criteria and appropriately extracted each study separately. Second, both studies have different designs

and methodologies, and obtain different outcomes. For example, one study is a co-designed study, and the other is a qualitative study; therefore, neither study should be included in the meta-analysis itself. Third, these studies are included and discussed in the review in relation to indicators of quality social connection within digital interventions and each study finding is different. In this case, different data are presented and referenced in relation to distinct digital intervention mechanisms that facilitate quality social connection and they are appropriately presented separately. For example, the importance of anonymity is mentioned in one but not the other. Finally, we

have already discussed the statistical and methodological variation within our scoping review as a potential limitation. Thank you for considering our response to the comment.

Acknowledgments

We are supported by Institute of Global Health Innovation, Imperial College London. This work was funded by a Wellcome Trust Mental Health Priority Area Active Ingredients commission awarded to LD at Imperial College London.

Conflicts of Interest

HA is the Chief Scientific Officer, Preemptive Medicine and Health Security at *Flagship Pioneering*.

References

1. Deng H, Qin X. Comment on 'Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis'. *J Med Internet Res* 2022:e36739 [FREE Full text] [doi: [10.2196/36739](https://doi.org/10.2196/36739)]
2. Dewa LH, Lawrance E, Roberts L, Brooks-Hall E, Ashrafian H, Fontana G, et al. Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis. *J Med Internet Res* 2021 Dec 17;23(12):e26584 [FREE Full text] [doi: [10.2196/26584](https://doi.org/10.2196/26584)] [Medline: [34927592](https://pubmed.ncbi.nlm.nih.gov/34927592/)]
3. Radovic A, DeMand AL, Gmelin T, Stein BD, Miller E. SOVA: Design of a stakeholder informed social media website for depressed adolescents and their parents. *J Technol Hum Serv* 2018;35(3):169-182 [FREE Full text] [doi: [10.1080/15228835.2017.1347552](https://doi.org/10.1080/15228835.2017.1347552)] [Medline: [29743822](https://pubmed.ncbi.nlm.nih.gov/29743822/)]
4. Radovic A, Gmelin T, Stein BD, Miller E. Depressed adolescents' positive and negative use of social media. *J Adolesc* 2017 Feb;55:5-15 [FREE Full text] [doi: [10.1016/j.adolescence.2016.12.002](https://doi.org/10.1016/j.adolescence.2016.12.002)] [Medline: [27997851](https://pubmed.ncbi.nlm.nih.gov/27997851/)]

Edited by T Leung; submitted 21.02.22; this is a non-peer-reviewed article; accepted 21.02.22; published 11.03.22.

Please cite as:

Dewa LH, Roberts L, Lawrance E, Ashrafian H

Authors' Reply to: Toward a Better Understanding of Quality Social Connections. Comment on "Quality Social Connection as an Active Ingredient in Digital Interventions for Young People With Depression and Anxiety: Systematic Scoping Review and Meta-analysis"
J Med Internet Res 2022;24(3):e37440

URL: <https://www.jmir.org/2022/3/e37440>

doi: [10.2196/37440](https://doi.org/10.2196/37440)

PMID: [35275072](https://pubmed.ncbi.nlm.nih.gov/35275072/)

©Lindsay H Dewa, Lily Roberts, Emma Lawrance, Hutan Ashrafian. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 11.03.2022. 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.

Letter to the Editor

Challenges in Measuring What Matters to Patients With Diabetes. Comment on “Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review”

Femke Rutters¹, PhD; Ellen Elsmann¹, PhD; Lenka Groeneveld¹, MSc; Marlous Langendoen-Gort¹, MSc; Lidwine Mokkink¹, PhD; Caroline Terwee¹, PhD

Department of Epidemiology and Data Science, Amsterdam University Medical Centre, Location Vrije Universiteit Amsterdam, Amsterdam, Netherlands

Corresponding Author:

Femke Rutters, PhD

Department of Epidemiology and Data Science

Amsterdam University Medical Centre, Location Vrije Universiteit Amsterdam

De Boelelaan 1089a

Amsterdam, 1081 HV

Netherlands

Phone: 31 204445860

Email: f.rutters@amsterdamumc.nl

Related Articles:

Comment on: <http://www.jmir.org/2021/8/e25002/>

Comment in: <http://mhealth.jmir.org/2022/8/e37957/>

(*J Med Internet Res* 2022;24(3):e36876) doi:[10.2196/36876](https://doi.org/10.2196/36876)

KEYWORDS

systematic review; measurement properties; patient-reported outcome measures; methodological quality; level of evidence; PROMs; patient reported outcome; diabetes

We would like to respond to Wee et al’s paper, “Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review” [1]. We appreciate the herculean effort undertaken to summarize all diabetes-related patient-reported outcome measures (PROMs). However, we have some concerns.

First, despite the large amount of identified PROMs (N=238), there are still many PROMs missing [1]. In our systematic review of PROMs measuring health-related quality of life (HRQOL) in people with type 2 diabetes (currently under review), which was performed in the same time period and using the same databases, we identified 116 HRQOL PROMS. Of these, >50 were missing in Wee et al’s review [1]. Missing PROMs include, for example, the National Diabetes Register Survey [2], which in our review showed the best content validity. We think this incompleteness is due to a lack of alternative search strategies, such as checking references. We were surprised that the authors [1] identified no papers through hand-searching, while about one-fourth of the included papers in our review were identified through reference checking.

Second, the authors [1] used the COSMIN (Consensus-Based Standards for the Selection of Health Measurement Instruments) methodology to summarize the evidence on the quality

(measurement properties) of the PROMs. However, contrary to the COSMIN guidelines, the quality of the PROMs was not rated for each PROM subscale separately, even though measurement properties can vary among subscales.

The limitations of this review [1] underscore the problematic status of PROMs in diabetes: there is no consensus on what doctors and scientists want to measure, and it is unclear what is most relevant to measure. The content of the existing PROMs is very heterogeneous; there are too many PROMs out there and many are of questionable validity. This hinders value-based health care and limits the value of PROMs when attempting to determine which treatment works most optimally. More awareness is needed, supported by recent initiatives on developing core outcome sets for people with diabetes [3-5]. We should start using those core outcome sets in our research and care for people with diabetes.

In conclusion, there is still a need for a high-quality systematic overview of all available PROMs for people with diabetes, with emphasis on the constructs being measured, and a comprehensive evidence synthesis of the measurement properties of all (subscales of) PROMs. This would allow researchers and doctors working with people with diabetes to

make informed choices on which PROMs to use for value-based health care.

Conflicts of Interest

None declared.

References

1. Wee PJJ, Kwan YH, Loh DHF, Phang JK, Puar TH, Ostbye T, et al. Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review. *J Med Internet Res* 2021 Aug 13;23(8):e25002 [FREE Full text] [doi: [10.2196/25002](https://doi.org/10.2196/25002)] [Medline: [34397387](https://pubmed.ncbi.nlm.nih.gov/34397387/)]
2. Svedbo Engström M, Leksell J, Johansson UB, Eeg-Olofsson K, Borg S, Palaszewski B, et al. A disease-specific questionnaire for measuring patient-reported outcomes and experiences in the Swedish National Diabetes Register: Development and evaluation of content validity, face validity, and test-retest reliability. *Patient Educ Couns* 2018 Jan;101(1):139-146. [doi: [10.1016/j.pec.2017.07.016](https://doi.org/10.1016/j.pec.2017.07.016)] [Medline: [28736071](https://pubmed.ncbi.nlm.nih.gov/28736071/)]
3. Dodd S, Harman N, Taske N, Minchin M, Tan T, Williamson PR. Core outcome sets through the healthcare ecosystem: the case of type 2 diabetes mellitus. *Trials* 2020 Jun 25;21(1):570 [FREE Full text] [doi: [10.1186/s13063-020-04403-1](https://doi.org/10.1186/s13063-020-04403-1)] [Medline: [32586349](https://pubmed.ncbi.nlm.nih.gov/32586349/)]
4. Harman NL, Wilding JPH, Curry D, Harris J, Logue J, Pemberton RJ, SCORE-IT Study Team. Selecting Core Outcomes for Randomised Effectiveness trials In Type 2 diabetes (SCORE-IT): a patient and healthcare professional consensus on a core outcome set for type 2 diabetes. *BMJ Open Diabetes Res Care* 2019 Dec;7(1):e000700 [FREE Full text] [doi: [10.1136/bmjdr-2019-000700](https://doi.org/10.1136/bmjdr-2019-000700)] [Medline: [31908789](https://pubmed.ncbi.nlm.nih.gov/31908789/)]
5. Patient-centered outcome measures: Diabetes. ICHOM Connect. 2019. URL: <https://connect.ichom.org/standard-sets/diabetes/> [accessed 2022-03-25]

Abbreviations

COSMIN: Consensus-Based Standards for the Selection of Health Measurement Instruments

HRQOL: health-related quality of life

PROM: patient-reported outcome measure

Edited by T Leung; submitted 31.01.22; this is a non-peer-reviewed article; accepted 21.03.22; published 31.03.22.

Please cite as:

Rutters F, Elsmann E, Groeneveld L, Langendoen-Gort M, Mokkink L, Terwee C

Challenges in Measuring What Matters to Patients With Diabetes. Comment on "Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review"

J Med Internet Res 2022;24(3):e36876

URL: <https://www.jmir.org/2022/3/e36876>

doi: [10.2196/36876](https://doi.org/10.2196/36876)

PMID: [35357322](https://pubmed.ncbi.nlm.nih.gov/35357322/)

©Femke Rutters, Ellen Elsmann, Lenka Groeneveld, Marlous Langendoen-Gort, Lidwine Mokkink, Caroline Terwee. Originally published in the *Journal of Medical Internet Research* (<https://www.jmir.org>), 31.03.2022. 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.

Letter to the Editor

Authors' Reply to: Challenges in Measuring What Matters to Patients With Diabetes. Comment on "Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review"

Yu Heng Kwan^{1,2,3}, BSc, MD, PhD; Jie Kie Phang⁴, BSc, MPH; Sungwon Yoon^{1,4}, MPH, PhD; Lian Leng Low^{1,4,5,6,7,8}, MBBS, MMed, MCI

¹Health Services and Systems Research, Duke-NUS Medical School, Singapore, Singapore

²SingHealth Internal Medicine Residency Programme, Singapore Health Services, Singapore, Singapore

³Department of Pharmacy, National University of Singapore, Singapore, Singapore

⁴Centre for Population Health Research and Implementation, SingHealth Regional Health System, Singapore, Singapore

⁵Population Health & Integrated Care Office (PHICO), Singapore General Hospital, Singapore, Singapore

⁶Department of Family Medicine and Continuing Care, Singapore General Hospital, Singapore, Singapore

⁷Post-Acute and Continuing Care, Outram Community Hospital, Singapore, Singapore

⁸SingHealth Duke-NUS Family Medicine Academic Clinical Program, Singapore, Singapore

Corresponding Author:

Yu Heng Kwan, BSc, MD, PhD

Health Services and Systems Research

Duke-NUS Medical School

8 College Rd

Singapore, 169857

Singapore

Phone: 65 90231226

Email: phakyh@nus.edu.sg

Related Articles:

Comment on: <http://www.jmir.org/2022/3/e36876/>

Comment on: <http://www.jmir.org/2021/8/e25002/>

(*J Med Internet Res* 2022;24(3):e37957) doi:[10.2196/37957](https://doi.org/10.2196/37957)

KEYWORDS

systematic review; measurement properties; patient-reported outcome measures; methodological quality; level of evidence; PROMs; patient reported outcome; diabetes

We would like to respond to the letter written by Rutters et al [1] with regard to our paper, "Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review" [2]. We noted the concerns from Rutters et al [1], but we would like to offer some explanations.

First, the selection criteria of our systematic review were restricted to patient-reported outcome measures (PROMs) that are tested in patients with type 2 diabetes mellitus (T2DM) only. The study on the development and validation of the National Diabetes Register Survey included patients with other forms of diabetes (ie, type 1 diabetes) [3], and, therefore, was excluded from our analysis. We focused on T2DM since existing evidence has demonstrated that patients' behaviors influencing disease

management differ by different diabetes subtypes [4,5]. Therefore, the PROMs used to guide interventions and patient care may be different and should be reviewed separately. Another consideration was related to our concerns that combining all validation studies of PROMs in different forms of diabetes would reduce the readability of the paper due to the large number of studies available.

Second, due to the large number of PROMs included in the review, we decided to analyze the measurement properties of the PROMs on a per-PROM basis instead to maintain the readability of the paper. We also agree with Rutters et al [1] that many health-related quality of life (HRQOL) PROM subscales do not measure HRQOL but actually measure overall

quality of life, and that characteristics of the individual or environment should be considered patient-reported experience measures. This is further complicated by the issue of problematic definitions of HRQOL in the literature [6]; thus, further study detailing the different constructs measured by subscales of PROMs is warranted.

We are grateful that the authors have taken the effort to provide constructive comments on our paper. The issues brought up by Rutters et al [1] echoed the need to have consensus between clinicians and psychometrists to measure what is relevant to patients. The content of the existing PROMs is indeed heterogeneous, and there are too many PROMs that have

questionable validity. We agree that more awareness is needed, including developing and implementing core outcome sets for patients with diabetes.

In conclusion, there is a need for a systematic review to summarize all available PROMs for patients with diabetes with emphasis on the constructs being measured, as well as a comprehensive evidence synthesis of the measurement properties of all subscales of PROMs (which was not the focus of our systematic review). Clinicians and researchers should work with patients with diabetes to develop a core outcome measurement set for use in diabetes care and research.

Conflicts of Interest

None declared.

References

1. Rutters F, Elsman E, Groeneveld L, Langendoen-Gort M, Mokkink L, Terwee C. Challenges in Measuring What Matters to Patients With Diabetes Comment on "Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review". *J Med Internet Res* 2022 Mar [FREE Full text] [doi: [10.2196/36876](https://doi.org/10.2196/36876)]
2. Wee PJL, Kwan YH, Loh DHF, Phang JK, Puar TH, Østbye T, et al. Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review. *J Med Internet Res* 2021 Aug 13;23(8):e25002 [FREE Full text] [doi: [10.2196/25002](https://doi.org/10.2196/25002)] [Medline: [34397387](https://pubmed.ncbi.nlm.nih.gov/34397387/)]
3. Engström MS, Leksell J, Johansson UB, Eeg-Olofsson K, Borg S, Palaszewski B, et al. A disease-specific questionnaire for measuring patient-reported outcomes and experiences in the Swedish National Diabetes Register: Development and evaluation of content validity, face validity, and test-retest reliability. *Patient Educ Couns* 2018 Jan;101(1):139-146. [doi: [10.1016/j.pec.2017.07.016](https://doi.org/10.1016/j.pec.2017.07.016)] [Medline: [28736071](https://pubmed.ncbi.nlm.nih.gov/28736071/)]
4. Trento M, Tomelini M, Basile M, Borgo E, Passera P, Miselli V, et al. The locus of control in patients with Type 1 and Type 2 diabetes managed by individual and group care. *Diabet Med* 2008 Jan;25(1):86-90. [doi: [10.1111/j.1464-5491.2007.02319.x](https://doi.org/10.1111/j.1464-5491.2007.02319.x)] [Medline: [18199135](https://pubmed.ncbi.nlm.nih.gov/18199135/)]
5. Sanz-Nogués C, Mustafa M, Burke H, O'Brien T, Coleman CM. Knowledge, Perceptions and Concerns of Diabetes -Associated Complications Among Individuals Living with Type 1 and Type 2 Diabetes Mellitus. *Healthcare (Basel)* 2020 Jan 30;8(1) [FREE Full text] [doi: [10.3390/healthcare8010025](https://doi.org/10.3390/healthcare8010025)] [Medline: [32019118](https://pubmed.ncbi.nlm.nih.gov/32019118/)]
6. Karimi M, Brazier J. Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? *Pharmacoeconomics* 2016 Jul;34(7):645-649. [doi: [10.1007/s40273-016-0389-9](https://doi.org/10.1007/s40273-016-0389-9)] [Medline: [26892973](https://pubmed.ncbi.nlm.nih.gov/26892973/)]

Abbreviations

HRQOL: health-related quality of life

PROM: patient-reported outcome measure

T2DM: type 2 diabetes mellitus

Edited by T Leung; submitted 13.03.22; this is a non-peer-reviewed article; accepted 21.03.22; published 31.03.22.

Please cite as:

Kwan YH, Phang JK, Yoon S, Low LL

Authors' Reply to: Challenges in Measuring What Matters to Patients With Diabetes. Comment on "Measurement Properties of Patient-Reported Outcome Measures for Diabetes: Systematic Review"

J Med Internet Res 2022;24(3):e37957

URL: <https://www.jmir.org/2022/3/e37957>

doi: [10.2196/37957](https://doi.org/10.2196/37957)

PMID: [35357327](https://pubmed.ncbi.nlm.nih.gov/35357327/)

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.

Corrigenda and Addenda

Correction: Using Participatory Design Methodologies to Co-Design and Culturally Adapt the Spanish Version of the Mental Health eClinic: Qualitative Study

Laura Ospina-Pinillos^{1,2}, MD; Tracey Davenport¹, BA (Hons), EMBA; Antonio Mendoza Diaz³, BPsych, PhD; Alvaro Navarro-Mancilla⁴, MPH, MD; Elizabeth M Scott⁵, MBBS, FRANZCP; Ian B Hickie¹, MD, FRANZCP

¹Brain and Mind Centre, The University of Sydney, Sydney, NSW, Australia

²Department of Psychiatry and Mental Health, Pontifical Javeriana University, Bogota, Colombia

³School of Psychiatry, Department of Medicine, University of New South Wales, Sydney, NSW, Australia

⁴Neuropsychiatry Research Group, Autonomous University of Bucaramanga, Bucaramanga, Colombia

⁵School of Medicine, University of Notre Dame Australia, Sydney, NSW, Australia

Corresponding Author:

Laura Ospina-Pinillos, MD

Brain and Mind Centre

The University of Sydney

Shops 1-3, 66-70 Parramatta Road

Sydney, NSW, 2051

Australia

Phone: 61 028 627 6946

Email: laura.ospinapinillos@sydney.edu.au

Related Article:

Correction of: <https://www.jmir.org/2019/8/e14127/>

(*J Med Internet Res* 2022;24(3):e37679) doi:[10.2196/37679](https://doi.org/10.2196/37679)

In “Using Participatory Design Methodologies to Co-Design and Culturally Adapt the Spanish Version of the Mental Health eClinic: Qualitative Study” (*J Med Internet Res* 2019;21(8):e14127), the authors noted one error.

In the originally published article, a value appeared incorrectly in the following sentence.

A total of 7 health professionals participated in the workshops; 6 were female and their ages ranged from 22 to 24 years (median age 28 years).

This has been corrected as follows:

A total of 7 health professionals participated in the workshops; 6 were female and their ages ranged from 22 to 34 years (median age 28 years).

The correction will appear in the online version of the paper on the JMIR Publications website on March 4, 2022, 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 02.03.22; this is a non-peer-reviewed article; accepted 02.03.22; published 03.03.22.

Please cite as:

Ospina-Pinillos L, Davenport T, Mendoza Diaz A, Navarro-Mancilla A, Scott EM, Hickie IB

Correction: Using Participatory Design Methodologies to Co-Design and Culturally Adapt the Spanish Version of the Mental Health eClinic: Qualitative Study

J Med Internet Res 2022;24(3):e37679

URL: <https://www.jmir.org/2022/3/e37679>

doi:[10.2196/37679](https://doi.org/10.2196/37679)

PMID:[35239500](https://pubmed.ncbi.nlm.nih.gov/35239500/)

©Laura Ospina-Pinillos, Tracey Davenport, Antonio Mendoza Diaz, Alvaro Navarro-Mancilla, Elizabeth M Scott, Ian B Hickie. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 03.03.2022. 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.

Corrigenda and Addenda

Correction: COVID-19 Vaccine Tweets After Vaccine Rollout: Sentiment–Based Topic Modeling

Luwen Huangfu^{1,2}, PhD; Yiwen Mo¹, MS; Peijie Zhang³, MS; Daniel Dajun Zeng^{3,4}, PhD; Saike He^{3,4}, PhD

¹Fowler College of Business, San Diego State University, San Diego, CA, United States

²Center for Human Dynamics in the Mobile Age, San Diego State University, San Diego, CA, United States

³The State Key Laboratory of Management and Control for Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing, China

⁴University of Chinese Academy of Sciences, Beijing, China

Corresponding Author:

Saike He, PhD

The State Key Laboratory of Management and Control for Complex Systems

Institute of Automation

Chinese Academy of Sciences

95 Zhongguancun East Road, Haidian District

Beijing, 100190

China

Phone: 86 (010)82544537

Email: saike.he@ia.ac.cn

Related Article:

Correction of: <https://www.jmir.org/2022/2/e31726>

(*J Med Internet Res* 2022;24(3):e37841) doi:[10.2196/37841](https://doi.org/10.2196/37841)

In “COVID-19 Vaccine Tweets After Vaccine Rollout: Sentiment–Based Topic Modeling” (*J Med Internet Res* 2022;24(2):e31726) the authors noted one error.

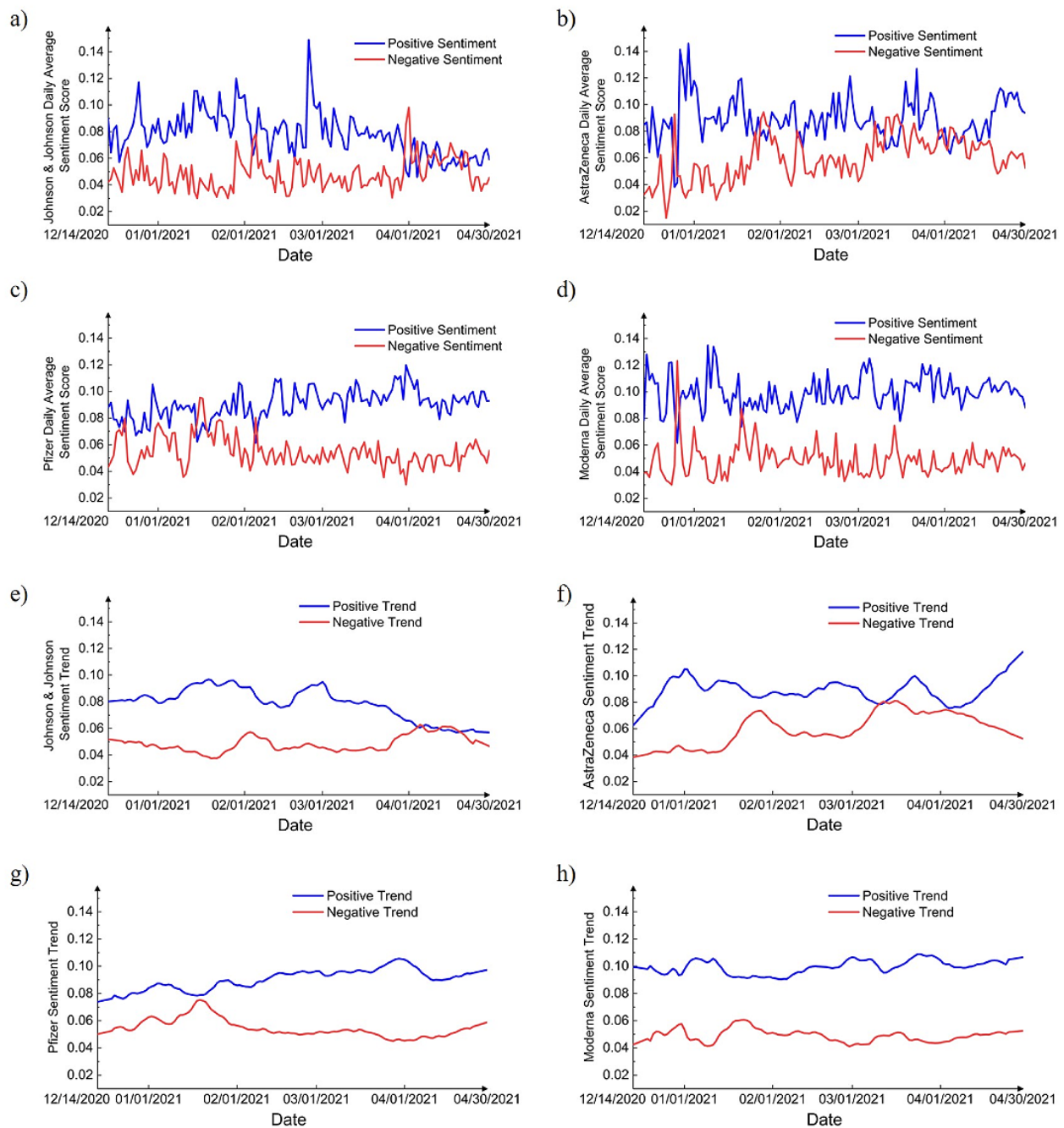
In the originally published paper, [Figure 8D](#) showed incorrect colors. The top line was intended to be blue, and the bottom line was intended to be red.

In the corrected version of the paper, [Figure 8D](#) has been revised as follows: the top line is now blue, and the bottom line is now

red. The correct figure is provided below. The originally published [Figure 8](#) is in [Multimedia Appendix 1](#).

The correction will appear in the online version of the paper on the JMIR Publications website on March 11, 2022, 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.

Figure 8. Daily average positive and negative sentiment scores for (a) Johnson & Johnson, (b) AstraZeneca, (c) Pfizer, and (d) Moderna vaccines and sentiment trends for (e) Johnson & Johnson, (f) AstraZeneca, (g) Pfizer, and (h) Moderna vaccines.



Multimedia Appendix 1

The previously published Figure 8.

[[PNG File , 502 KB - jmir_v24i3e37841_app1.png](#)]

Submitted 08.03.22; this is a non-peer-reviewed article; accepted 09.03.22; published 11.03.22.

Please cite as:

Huangfu L, Mo Y, Zhang P, Zeng DD, He S

Correction: COVID-19 Vaccine Tweets After Vaccine Rollout: Sentiment-Based Topic Modeling

J Med Internet Res 2022;24(3):e37841

URL: <https://www.jmir.org/2022/3/e37841>

doi: [10.2196/37841](https://doi.org/10.2196/37841)

PMID: [35275838](https://pubmed.ncbi.nlm.nih.gov/35275838/)

©Luwen Huangfu, Yiwen Mo, Peijie Zhang, Daniel Dajun Zeng, Saike He. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 11.03.2022. 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.

Corrigenda and Addenda

Metadata Correction: Clinical Outcomes Among Working Adults Using the Health Integrator Smartphone App: Analyses of Prespecified Secondary Outcomes in a Randomized Controlled Trial

Stephanie Bonn¹, PhD; Gabriella Licitra¹, MD; Rino Bellocco^{2,3}, PhD; Ylva Trolle Lagerros^{1,4}, MD, PhD

¹Clinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

²Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

³Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy

⁴Center for Obesity, Academic Specialist Center, Stockholm Health Services, Stockholm, Sweden

Corresponding Author:

Stephanie Bonn, PhD

Clinical Epidemiology Division

Department of Medicine Solna

Karolinska Institutet

T2, Karolinska University Hospital

Stockholm, 171 76

Sweden

Phone: 46 08 51779173

Email: stephanie.bonn@ki.se

Related Article:

Correction of: <https://www.jmir.org/2022/3/e24725>

(*J Med Internet Res* 2022;24(3):e38199) doi:[10.2196/38199](https://doi.org/10.2196/38199)

In “Clinical Outcomes Among Working Adults Using the Health Integrator Smartphone App: Analyses of Prespecified Secondary Outcomes in a Randomized Controlled Trial” (*J Med Internet Res* 2022;24(3):e24725), the following correction was made:

In the originally published article, the names of peer reviewers’ names inadvertently included an additional reviewer.

The correct statement of peer review should be “*peer-reviewed by X Guo.*”

The correction will appear in the online version of the paper on the JMIR Publications website on March 24, 2022, 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 22.03.22; this is a non-peer-reviewed article; accepted 22.03.22; published 24.03.22.

Please cite as:

Bonn S, Licitra G, Bellocco R, Trolle Lagerros Y

Metadata Correction: Clinical Outcomes Among Working Adults Using the Health Integrator Smartphone App: Analyses of Prespecified Secondary Outcomes in a Randomized Controlled Trial

J Med Internet Res 2022;24(3):e38199

URL: <https://www.jmir.org/2022/3/e38199>

doi: [10.2196/38199](https://doi.org/10.2196/38199)

PMID: [35324452](https://pubmed.ncbi.nlm.nih.gov/35324452/)

©Stephanie Bonn, Gabriella Licitra, Rino Bellocco, Ylva Trolle Lagerros. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 24.03.2022. 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.

Corrigenda and Addenda

Correction: Gender-Specific Impact of Self-Monitoring and Social Norm Information on Walking Behavior Among Chinese College Students Assessed Using WeChat: Longitudinal Tracking Study

Yuepei Xu^{1,2*}, BSc; Ling-Zi Yue^{1,2*}, MSc; Wei Wang³, PhD; Xiao-Ju Wu^{1,2}, MSc; Zhu-Yuan Liang^{1,2}, PhD

¹CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, China

²Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

³Brigham and Women's Hospital, Harvard Medical School, Boston, MA, United States

*these authors contributed equally

Corresponding Author:

Zhu-Yuan Liang, PhD

CAS Key Laboratory of Behavioral Science

Institute of Psychology

16 Lincui Road, Chaoyang District

Beijing, 100101

China

Phone: 86 10 64841536

Email: liangzy@psych.ac.cn

Related Article:

Correction of: <https://www.jmir.org/2021/12/e29167>

(*J Med Internet Res* 2022;24(3):e38221) doi:[10.2196/38221](https://doi.org/10.2196/38221)

In “Gender-Specific Impact of Self-Monitoring and Social Norm Information on Walking Behavior Among Chinese College Students Assessed Using WeChat: Longitudinal Tracking Study” (*J Med Internet Res* 2021;23(12):e29167), one error was noted.

The foundation number of the National Natural Science Foundation of China was mistaken. In the originally published paper, under “Acknowledgments”, the foundation information was listed as follows:

This research was supported by Beijing Natural Science Foundation (BNSF, 9172019), the National Natural Science Foundation of China (NSFC, 7170111) ...

This has been corrected to:

This research was supported by Beijing Natural Science Foundation (BNSF, 9172019), the National Natural Science Foundation of China (NSFC, 71471171) ...

The correction will appear in the online version of the paper on the JMIR Publications website on March 31, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 24.03.22; this is a non-peer-reviewed article; accepted 24.03.22; published 31.03.22.

Please cite as:

Xu Y, Yue LZ, Wang W, Wu XJ, Liang ZY

Correction: Gender-Specific Impact of Self-Monitoring and Social Norm Information on Walking Behavior Among Chinese College Students Assessed Using WeChat: Longitudinal Tracking Study

J Med Internet Res 2022;24(3):e38221

URL: <https://www.jmir.org/2022/3/e38221>

doi:[10.2196/38221](https://doi.org/10.2196/38221)

PMID:[35358056](https://pubmed.ncbi.nlm.nih.gov/35358056/)

©Yuepei Xu, Ling-Zi Yue, Wei Wang, Xiao-Ju Wu, Zhu-Yuan Liang. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 31.03.2022. 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.

Original Paper

Influence of Forced Online Distance Education During the COVID-19 Pandemic on the Perceived Stress of Postsecondary Students: Cross-sectional Study

Andrej Šorgo¹, PhD; Nuša Crnkovič², MSc; Branko Gabrovec², PhD; Katarina Cesar², MSc; Špela Selak², PhD

¹Faculty of Natural Science and Mathematics, University of Maribor, Maribor, Slovenia

²National Institute of Public Health, Ljubljana, Slovenia

Corresponding Author:

Špela Selak, PhD

National Institute of Public Health

Trubarjeva cesta 2

Ljubljana, 1000

Slovenia

Phone: 386 1 620 36 45

Email: spela.selak@nijz.si

Abstract

Background: One of the most significant changes in the majority of postsecondary educational institutions was the closure of those institutions and the shift of educational activities to online distance learning formats as a result of the COVID-19 pandemic. Closure combined with forced online distance education (FODE) was a cure with many side effects, 1 of them being the effect on students' mental health and, more specifically, levels of stress. Due to the novelty of the situation, there have been no studies so far designed to link satisfaction with online study, feelings toward the study obligations, and stress among students.

Objective: The aim of the study is to assess the perceived stress of Slovenian postsecondary students in order to identify the online study-related factors affecting or acting as a covariate during the COVID-19 lockdown.

Methods: Data collection was conducted through a self-reported survey as part of a large cross-sectional study based on data collected from postsecondary students from a number of higher educational institutions. The random sample consisted of 4455 individuals. The Perceived Stress Scale (PSS-4), Satisfaction with Online Study Scale (SAT-5), and Feelings Towards Study Obligations Scale (FETSOS) were used to assess the constructs and the relations observed within the study.

Results: The results indicate that more than half of all respondents reported high levels of stress. The difference in the reported levels of perceived stress between genders were statistically significant ($N=4454$, $F_2=56.719$, $P<.001$, Cohen $d=0.35$). Overall, the results suggest that a decline in the motivation to study, the quality of internet and mobile connections, and the presence of distracting factors in the study space were the 3 main factors related to the students' negative emotions as associated with the timeliness, performance, and quality of the study obligations. Furthermore, the results show that the level of satisfaction with online study affected stress such that the higher the satisfaction, the lower the stress. Moreover, the more positive feelings connected with the timeliness, performance, and quality of the study obligations that the students felt, the more satisfaction they reported with online study and, thus indirectly, lower stress and less negative feelings.

Conclusions: The findings of this study call for implementing structures and measures targeted at stress reduction, working conditions, and pedagogy with regard to FODE.

(*J Med Internet Res* 2022;24(3):e30778) doi:[10.2196/30778](https://doi.org/10.2196/30778)

KEYWORDS

online study; stress; COVID-19; postsecondary students; pandemic; epidemiology; educational institutions; online education; pedagogy; mental health

Introduction

Academic environments have always been known for the presence of stress, anxiety, and depression [1,2]. The closure of postsecondary educational institutions induced by COVID-19 brought about many changes to almost all areas of study. In this context, study mostly takes place online and at a distance. Because of its compulsory nature, it has been recognized as forced distance education (teaching, learning) and in its online form as forced online distance education (FODE) [3]. Moving lectures and courses online has not only changed the format of courses but also brought with it several side effects [4]. One of the most often reported side effects of FODE is the impact on students' mental health, with stress, anxiety, and depression being the most commonly reported impacts worldwide [5-7], with some subpopulations being more vulnerable than others [8]. Students seem to be at greater risk of mental health problems due to social distancing, as well as FODE and other measures to prevent or reduce the risk of transmission of coronavirus (SARS-CoV-2), such as restrictions on movement and gatherings [9]. Previous research has shown a link between psychological distress and symptoms of mental disorders or problems with outbreaks of infectious diseases [10,11]. Research has shown that during the COVID-19 pandemic also, psychological reactions, such as stress, anxiety, and depression, have been common [12,13] among students as well [14-21]. Several studies have reported high or increased prevalence of stress among this population [6,8,17,19,22,23], associated with factors such as fear and worry about one's own health or that of loved ones, difficulties concentrating, disrupted sleeping patterns, reduced social interaction, and increased concerns about academic performance [19], which raises concerns about the mental health of postsecondary students.

The studies reviewed most often searched for predictors of increased mental health problems during a lockdown in socioeconomic domains, previous (mental) health episodes, living conditions, and, more rarely, the domain of education [24-26]. The majority of studies, produced by professionals from the fields of psychology and psychiatry, most often suggest solutions for building resilience among students, counselling, and, at the last instance, medication [8,27]. However, the solution for students can lie not only in improving self-supportive and institutionally supported measures for those with mental health problems but also in reducing sources of stress where and when possible. However, many studies of online distant (remote) education in the educational research domain are from the pre-COVID-19 era. The pandemic has put these previous findings into a new context. Those pre-COVID-19 findings cannot be easily transferred, because the earlier studies were based mostly on voluntary decisions, while at least during the first lockdown, the transition was forced and not tailored to the best standards of remote education. In the words of Hodges et al [28], "Everyone involved in this abrupt migration to online learning must realize that these crises and disasters also create disruptions to student, staff, and faculty lives, outside their association with the university."

In recent research, perceived stress has been measured as an output variable, as a covariate with several manifestations of

mental health problems, such as anxiety and depression (among numerous others), commonly associated with pandemic measures [12,29]. The input constructs included factors extracted from the exploratory analysis of variables believed to be stressors produced in relation to learning experiences, and thus can be manipulated by educational institutions and educators. As an intellectual framework, we followed the Sternberg theory of successful intelligence [30], as elaborated for FODE in Dolenc et al [4], which claims that educators "have the choice of adapting to the new environment, adapting the environment or changing the environment, while students can only adapt to the environment."

Postsecondary students can be regarded as 1 of the most important investments of every society in its future prosperity [1]. It is obvious that the share of individuals with higher education should and will rise in the coming generations [31]. There is also a trend toward the digital transformation of education, with the transition to more or less blended forms of education with promises of "anywhere (any place) – any time" learning experiences [32,33]. These trends call not only for an increase in the number of study places and the digital transformation of communication channels at educational institutions (colleges, higher schools, academies, faculties, universities, etc) but also for greater support for student well-being. In the transitional period between adolescence and adulthood, postsecondary educational institutions should be concerned with the quality of the learning outcomes of diverse student populations in line with student health and well-being. As such, the findings and experiences from the COVID-19 lockdowns should not be forgotten but should be taken not only as an opportunity to identify problems that may have been masked and unacknowledged before the lockdown but also as an incentive to address them. In line with these intentions, the first part of this study is descriptive, with the second part suggesting prescriptive measures.

The aim of the study was twofold. The primary focus was on assessing the mental health of Slovenian postsecondary students and identifying the online study-related factors affecting or acting as a covariate with it during the COVID-19 lockdown. The data used in this study were collected from postsecondary students from various higher educational institutions as part of a project titled "Measures in the Field of COVID-19 Spread Management With a Focus on Vulnerable Populations" [34]. The assumptions of the study were that technology was not neutral and that satisfaction with online experiences and with forced and involuntary study conditions during the lockdown could work as an incubator to raise perceived stress, as shown in previous studies [34]. The authors are aware that a number of factors and variables may correlate with, be influenced by, or predict stress. Among others, we can list (in no particular order) anxiety, depression, resilience, fear of COVID-19, previous mental illness, substance abuse, and addiction. However, at this point, such relationships were not explored. This was not because they were unimportant but because the primary goal was to explore the effects of FODE on stress within a simple and robust model.

The study was divided into 2 parts, in line with the research questions (RQs):

- RQ1: What are the latent structure, reliability, and construct validity of the instruments used in the survey?
- RQ2: What are the strength and direction of paths between latent variables?

The study results may help us produce prescriptive measures to help improve study conditions and the mental health of postsecondary students and also help those who may be concerned with the problem and have the ability and means to take action during and after the COVID-19 pandemic.

Methods

Procedures and Instruments

Data collection was conducted through a self-reported survey as part of a large cross-sectional study in Slovenia. The data collection took place between February 9 and March 8, 2021. Data collection was conducted through the web-based survey platform 1KA (Centre for Social Informatics, at the Faculty of Social Sciences, University of Ljubljana) [35]. Simple random sampling was used, and invitation letters to participate in the study were sent by email to all universities, private faculties, and student organizations, with a request to forward the invitation to all their students. To obtain as much feedback as possible, a reminder letter with the invitation to participate was sent to all addresses after 1 week and, after another week, to those from whom we had not received any feedback. Respondents were informed about the various aspects of the study, including their right to voluntarily participate in and withdraw from it.

An online questionnaire (see [Multimedia Appendix 1](#)) was used for data collection. In addition to demographic data, metric tools were used to assess perceived stress, satisfaction with online study, and feelings toward the study obligations.

The Slovenian translation of the Perceived Stress Scale (PSS-4; see [Multimedia Appendix 2](#)), a shortened version of PSS-10 and PSS-14, was used for the assessment of psychological stress [36-38]. The items in the instrument ask respondents to report on how they have coped with various situations over the past month. The instrument has 4 items, 2 of which (items STR2 and STR3) have a reverse score. The response format is 0=never; 1=almost never; 2=sometimes; 3=fairly often; and 4=very often, with higher totals indicating higher levels of stress (the scores range from 0 to 16). The perceived stress level was categorized as low versus high perceived stress based on a median split (<8 vs ≥8) in analogy with application of the PSS-10 instrument in the Slovenian sample [39].

The Satisfaction with Online Study Scale (SAT-5; see [Multimedia Appendix 3](#)), initially developed by Debevc et al [40] and adapted by Ploj-Virtič et al [3], was used for assessment of satisfaction with online study. The scale is rooted in flow theory [41] and was applied in slightly different versions [3,42]. The measurement encompassed a 7-point Likert scale ranging from 1=strongly disagree to 7=strongly agree, with a total score ranging from 5 to 35 and higher scores indicating higher perceived satisfaction. The scale shows unidimensionality and Cronbach α values >.80 in all studies in which it was used.

The severity of several factors inducing negative feelings connected with the timeliness, performance, and quality of the study obligations were assessed using the Feelings Toward Study Obligations Scale (FETSOS; see [Multimedia Appendix 4](#)). The scale was designed for the purpose of this study. The authors consulted the outcomes of the study by Dolenc et al [4] and transferred some of their findings into statements (items) of the scale. The response format was 1=no impact; 2=very weak impact; 3=weak impact; 4=moderate impact; 5=strong impact; 6=very strong impact; and 7=absolute impact. The scale has 12 items. Because of their diversity, it was not expected to be unidimensional, which was later confirmed. Theoretically, the span of the scale is between 12 and 84, with higher numbers indicating more negative impacts. Three subscales can be identified: descriptors of working conditions, descriptors of pedagogy, and descriptors of well-being and health.

Quantitative measures of all subscales are provided in the Multimedia Appendices.

Ethics Approval

Ethical approval to conduct the study was obtained from the National Medical Ethics Committee of the Republic of Slovenia (NMEC), Ministry of Health (no. 0120-48/2021/3).

Statistical Analyses

We collected responses from 5999 full-time students. Due to the planned analyses, attrition, and random missing responses, the database was cleared and only data sets from those individuals who provided full responses for all the constructs were selected (4455/5999, 74.26%). We calculated the frequencies and measures of central tendencies for each item and the sums of items, when appropriate. Factor analysis was conducted to examine and validate the factor structure of the data matrix of the scales. In the factor analysis, the total sample (N=4455) was randomly divided into an exploratory factor analysis (EFA) sample (n=2235, 50.17%) and a confirmatory factor analysis (CFA) sample (n=2220, 49.83%). EFA was conducted to examine the underlying factor structure of the constructs in the EFA sample using principal axis factoring (PAF) analysis. Because correlations between components were expected, direct oblimin rotations were applied. Parallel analysis was used to determine the number of factors extracted from the EFA [43]. Cronbach α >.70 and unidimensionality of the construct were the entrance criteria to be included in CFA.

Two approaches were used for building the models. The first was based on correlations between the sums of the extracted constructs, where a path direction was not guessed. Spearman ρ was used for this. The second model was based on structural equation modeling (SEM) analysis [44,45], where hypothesized models were tested to fit the data. The maximum likelihood method was used, and analysis of the residual covariance matrix and inspection of the modification indices were applied to improve model fits. The whole data set (N=4455) was included in the final SEM analyses. We chose a selection of fit indexes, as proposed by Gaskin and Lim [46], and applied it using the IBM AMOS plugin proposed by Hu and Bentler [47]. The cut-off criteria recognized as acceptable are as follows: comparative fit index (CFI)>0.90, standardized

root-mean-square residual (SRMR) <0.08 , and root-mean-square error of approximation (RMSEA) <0.06 . CFA with maximum likelihood estimation [48] was used for model fitting with the application of IBM AMOS 27 and IBM SPSS Statistics 27 for EFA.

Research Model

The hypothesized research models were based on the following hypothesized paths:

- Hypothesis 1 (H1): Level of satisfaction with online study influences (correlates with) stress.
- H2: Feelings toward online study (working conditions, pedagogy, and well-being) influence (correlate with) satisfaction.
- H3: Feelings toward online study (working conditions, pedagogy, and well-being) influence (correlate with) stress.

It was assumed that satisfaction would work as a mediator between feelings toward online study (working conditions, pedagogy, and well-being) and stress (Figure 1).

Figure 1. Framework of the study.



Results

Characteristics of the Survey Respondents

The demographic characteristics of the students (Table 1) included in the study show there were more males (n=3234, 72.59%) than females included in the sample and more than

half of the students were single (n=2467, 55.38%). The majority were bachelor's degree students (n=2696, 60.52%). The most numerous study fields were health and medicine (n=886, 19.89%), science and mathematics (n=871, 19.55%), noneducational social studies (n=801, 17.98%), and humanities (n=707, 15.87%).

Table 1. Sample characteristics (N=4455).

Demographic characteristics	Frequency, n (%)
Gender	
Male	3234 (72.59)
Female	1186 (26.62)
Other	34 (0.76)
Missing	1 (0.02)
Educational level	
Higher vocational	129 (2.90)
Bachelor's study	2696 (60.52)
Master's study	1606 (36.05)
Doctoral study	18 (0.40)
Other	6 (0.13)
Relationship	
Single	2467 (55.38)
In a relationship	1919 (43.08)
Other	68 (1.53)
Missing	1 (0.02)
Field of study	
Health and medicine	886 (19.89)
Science and mathematics	871 (19.55)
Social studies (noneducation)	801 (17.98)
Humanities	707 (15.87)
Art	409 (9.18)
Technology and engineering	369 (8.28)
Education	251 (5.63)
Security	128 (2.87)
Other	33 (0.74)

Latent Structure, Reliability, and Construct Validity of the Instruments

EFA of PSS-4 (n=2235, 50.17%) revealed that it is a unidimensional factor (latent variable), with Cronbach $\alpha=.80$. The explained variance was 50.3% (eigenvalue=2.496). Measures of the central tendencies of 2 positively (reversed) and 2 negatively worded items, as well of the results of PAF, can be seen in [Multimedia Appendix 2](#).

CFA (n=2220, 49.83%) was performed with 1- and 2-factor models. The difference is that in the 1-factor model, all 4 items load on a single factor, whereas in the 2-factor model, the positively loaded items load on the first factor and the negatively loaded items load on the second factor. A 1-factor solution, even if the 2-factor model shows a slightly better fit, was chosen to be included in the models used to predict the hypotheses. The reason was that we wished to include PSS-4 as a complete and valid instrument. The values of the FIT indexes with constrained STR2 and STR3 ($r=0.38$) were as follows: chi-square

(CMIN)=31.824, $df=1$, CMIN/ $df=31.824$, CFI=0.995, SRMR=0.013, and RMSEA=0.081.

According to EFA, SAT-5 is a unidimensional tool (Cronbach $\alpha=.88$), and the first factor (eigenvalue=3.399) explained 67.971% of the variance. The values of the FIT indexes for the SAT-5 were as follows: CMIN=36.181, $df=3$, CMIN/ $df=12.060$, CFI=0.995, SRMR=0.016, and RMSEA=0.071. Constrained pairs of error terms were between SAT2 and SAT5 ($r=-0.36$) and SAT3 and SAT4 ($r=-0.22$). Measures of central tendencies, communalities, and factor loadings of SAT-5 are presented in [Multimedia Appendix 3](#).

Analysis of the results obtained by FETSOS (Cronbach $\alpha=.88$) revealed a 2-factor structure, cumulatively explaining 56.30% of the variance. The first factor included all 12 listed items (eigenvalue=5.295, variance=44.126%, Cronbach $\alpha=.88$), while the second factor consisted of 2 items (equipment, and mobile and internet connections) negatively cross-loading to the first factor as well (eigenvalue=1.461, variance=12.173, Cronbach $\alpha=.90$; see [Multimedia Appendix 4](#)). The structure obtained did

not follow the theoretically predicted subscales (pedagogy, working conditions, and well-being; see [Multimedia Appendix 4](#)). Measures of central tendencies, communalities, and factor loadings of FETSOS are presented in [Multimedia Appendix 4](#).

Differences of Perceived Stress, Satisfaction With Online Study, and Feelings Toward Study Obligations Between Genders and Study Enrolment Levels

Analysis of perceived stress level categorized as low versus high perceived stress based on a median split (<8 vs ≥8; N=4455) revealed that the sample mean was 7.97 (SD 3.32) and that 1938 (43.50%) of the 4455 students belonged to the lower-stress group and more than half of them to the high-stress group. The upper quarter (≥13) contained 412 (9.25%) of the 4455 students.

The difference in the means of the reported levels of perceived stress ([Table 2](#)) between genders were statistically significant, with the highest for those who reported a nonbinary gender, followed by females, both above the median split. Males were

below the median split. The effect size between males and females could be regarded as a small effect.

When comparing students from the 2 educational levels ([Table 3](#)), it appeared that those studying for a bachelor’s degree experienced a higher level of stress than those studying for a master’s degree or a doctorate. The effect size between both levels, calculated as Cohen *d*, could be regarded as a small effect.

Regarding satisfaction with online study, an examination of the measures of central tendencies (see [Multimedia Appendix 3](#)) revealed that most of the students thought that their online experiences were comprehensible, successful, and instructive but not easy or entertaining.

The difference between genders based on the means of the responses to SAT-5 was not statistically significant (N=4454, $F_2=0.18, P=.98$). Therefore, the results are not reported in the table.

Table 2. Differences between genders for PSS-4^a and FETSOS^b scores (N=4454).

Gender	PSS-4 ($F_2=56.719, P<.001, \text{Cohen } d=0.35$ [95% CI=0.284-0.418])			FETSOS ($F_2=74.771, P<.001, \text{Cohen } d=-0.412$ [95% CI=0.345-0.479])		
	Mean (SD)	95% CI		Mean (SD)	95% CI	
Men	7.12 (3.31)	6.93-7.30		49.18 (15.32)	48.30-50.05	
Women	8.27 (3.26)	8.16-8.39		55.13 (14.109)	54.65-55.62	
Nonbinary	9.21 (3.18)	8.10-10.32		57.03 (15.15)	51.74-62.32	

^aPSS-4: Perceived Stress Scale.

^bFETSOS: Feelings Towards Study Obligations Scale.

Table 3. Differences between educational levels for PSS-4^a, SAT-5^b, and FETSOS^c scores (N=4467).

Educational level	PSS-4 ($F_1=27.201, P<.001, \text{Cohen } d=0.164$ [95% CI=-0.103 to 0.225])			SAT-5 ($F_1=4.231, P=.04, \text{Cohen } d=0.064$ [95% CI=0.003-0.125])			FETSOS ($F_1=9.889, P=.002, \text{Cohen } d=0.097$ [95% CI=0.037-0.158])		
	n (%)	Mean (SD)	95% CI	n (%)	Mean (SD)	95% CI	n (%)	Mean (SD)	95% CI
Bachelor’s study	2825 (63.24)	8.17 (3.31)	8.05-8.29	2825 (63.24)	17.77 (8.02)	17.48-18.07	2825 (63.24)	54.08 (14.791)	53.54-54.63
Master’s or doctoral study	1642 (36.76)	7.63 (3.26)	7.47-7.79	1642 (36.76)	18.27 (7.43)	17.91-18.46	1642 (36.76)	52.65 (14.458)	51.94-53.35

^aPSS-4: Perceived Stress Scale.

^bSAT-5: Satisfaction with Online Study Scale.

^cFETSOS: Feelings Towards Study Obligations Scale.

When comparing students from the 2 educational levels, it appeared that those from the bachelor’s level were less satisfied than students from the master’s or doctoral levels ([Table 3](#)). The effect size between both levels was negligible, which can, in practice, be regarded as the absence of an effect. Furthermore, regarding feelings toward the study obligations, according to the opinions of the students (see [Multimedia Appendix 4](#)), the top 3 items that negatively influenced their feelings connected with the timeliness, performance, and quality of the study obligations are (1) a decline in the motivation to study, (2) quality of internet and mobile connections, and (3) the presence of distractions in the study space (eg, other people). In contrast,

the 3 items with the least negative impact were (1) the need to earn an income, (2) health problems directly related to distance learning, and (3) health problems not directly related to distance learning.

The difference between genders based on the means of the responses were statistically significant, with the highest for those who reported a nonbinary gender, followed by females and males ([Table 2](#)). From the SD values, a huge variation in the responses could be observed. The effect size between males and females, calculated as Cohen *d*, could be regarded as a small effect.

When comparing students from 2 educational levels, it appeared that those from the bachelor's level reported overall higher negative feelings than students from the master's or doctoral levels (Table 3). The effect size between both levels, calculated as Cohen *d*, could be regarded as a small effect.

Strength and Direction of Paths Between Latent Variables

Correlation analysis between sums of scales (Table 4) showed that negative feelings toward online study (FETSOS) moderately and negatively correlated with (SAT-5) and moderately correlated with stress (PSS-4).

Furthermore, the results of SEM analysis with SAT as a moderator variable (Figure 2) revealed that 18% of the variance of SAT-5 could be explained by FETSOS and 17% of the variance of stress as the main outcome variable could be explained by the joint influence of FETSOS and SAT-5. The

values explaining model fits were CFI=1, SRMR=0, and RMSEA=0.359, which led us to build alternative models.

The hypothetical model followed all 3 proposed hypotheses, and all variables were included without constraints (Figure 3). It was found that the model fit measures were outside the acceptable levels (CMIN=10380.95, *df*=186, CMIN/*df*=55.812, CFI=0.779, SRMR=0.097, and RMSEA=0.111). We therefore started procedures to improve the model fit. Because we did not want to delete items, as proposed by an examination of the standardized residual covariances in order to preserve the established scales, we applied constraint of error variances. We ended the procedure when acceptable fits were achieved (CMIN=4068.80, *df*=179, CMIN/*df*=22.731, CFI=0.916, SRMR=0.076, and RMSEA=0.070).

The values of the regression coefficients followed the same pattern as the correlational and SEM analysis of sums, but all the values were somewhat higher. We could explain 25% of the stress using the proposed model.

Table 4. Correlations^a (Spearman ρ) between totals for PSS-4^b, SAT-5^c, and FETSOS^d (N=4555).

	FETSOS	SAT-5	PSS-4
FETSOS	— ^e	—	—
SAT-5	-.390	—	—
PSS-4	.316	-.361	—

^aAll correlations were significant at the 0.01 level (2-tailed).

^bPSS-4: Perceived Stress Scale.

^cSAT-5: Satisfaction with Online Study Scale.

^dFETSOS: Feelings Towards Study Obligations Scale.

^eNot applicable.

Figure 2. Path coefficients between sums of scales. FETSOS: Feelings Towards Study Obligations Scale; PSS-4: Perceived Stress Scale; SAT-5: Satisfaction with Online Study Scale.

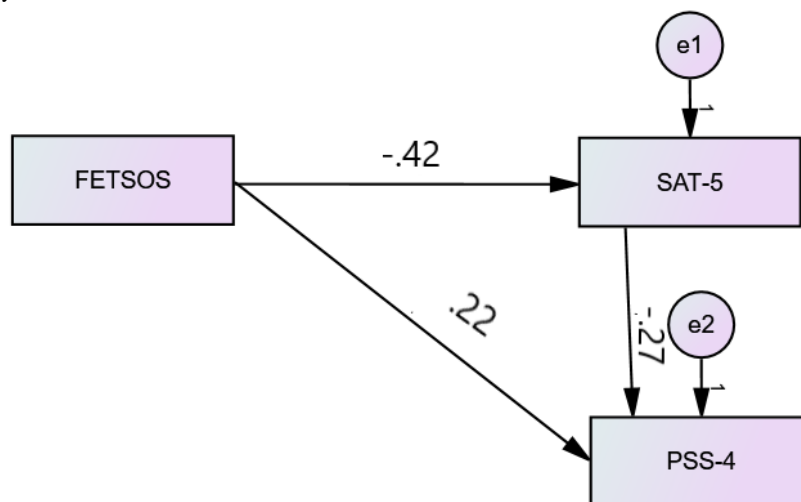
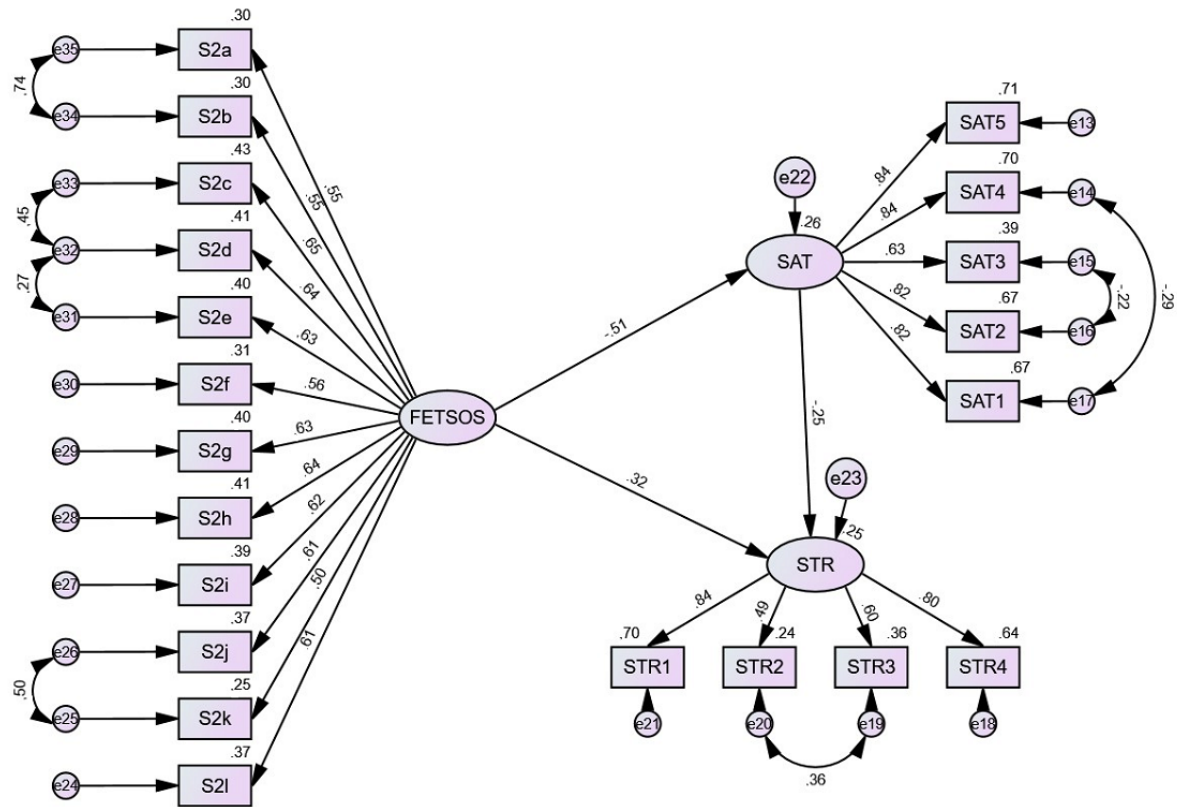


Figure 3. Measurement and structural model connecting FETSOS (variables S2a-S2l), SAT-5, and STR (PSS-4). FETSOS: Feelings Towards Study Obligations Scale; PSS-4: Perceived Stress Scale; SAT-5: Satisfaction with Online Study Scale.



Discussion

Principal Findings

The long-lasting COVID-19 pandemic, with its numerous accompanying preventive measures, imposed a variety of challenges affecting the mental health of the population, with young and emerging adults, which coincides with the beginning of postsecondary education, being more significantly affected [49]. This group already faces a number of stress-related and mental health difficulties due to the characteristics of this transitional period of life, such as instability owing to changes in education, living arrangements, and relationships. COVID-19-related challenges might additionally impact the latter [49,50]. The aim of this study was therefore to examine the levels of stress of Slovenian postsecondary students during the forced online distance learning occasioned by COVID-19 lockdowns and to identify the factors influencing it.

In response to the pandemic, many countries replaced face-to-face education with distance education. This could potentially result in negative social, psychological, and academic consequences for postsecondary students [51]. The results of this study indicate that more than half of the students reported higher levels of stress, with students studying for bachelor's degrees reporting higher levels than those studying for a master's degree or a doctorate. This is in line with previous research, which showed higher levels of stress among students engaged in distance education [5,6,17-20,22,23,52-54]. Stress levels also differed significantly between gender groups, with females reporting higher levels of perceived stress than males. This could be attributed to the greater vulnerability of women to the

development of mental health problems in general [55,56], as well during the COVID-19 pandemic [18,20,49,51,57].

Although the students who participated in our study reported the online study experience as being comprehensible, successful, and instructive, the overall satisfaction with online study was the lowest among bachelor's degree students, with no significant differences between the gender groups. Overall, the results show 3 main factors to be related to students' negative emotions associated with the timeliness, performance, and quality of the study obligations: a decline in the motivation to study, the quality of internet and mobile connections, and the presence of distractors in the study space (eg, other people). Once again, bachelor's degree students reported higher levels of negative emotions associated with the aforementioned factors compared to other student groups. These findings are in line with previous reports, which show COVID-19 to have had a negative impact on the academic experiences of postsecondary students [49,58,59], with distance education resulting in higher levels of stress and isolation; a negative mood; and lower levels of relatedness, concentration, focus, motivation, and performance compared to face-to-face education [58].

All 3 initial hypotheses were tested and supported. We were able to explain 25% of stress measured by PSS-4 with the SEM full model and 17% by use of the sums of the responses. The results from correlational and regression analyses showed that the level of satisfaction with online study influences stress (H1) in such a way that the higher the satisfaction, the lower the stress (Spearman $\rho = -.361$, path coefficient = $-.25$). These results are in line with Lee and Jang's findings [60], indicating a negative correlation between students' overall stress and satisfaction

scores with their study life, although their research was not specifically focused on online study.

The second hypothesis, that feelings toward online study (working conditions, pedagogy, and well-being) influence satisfaction, was also confirmed. The results (Spearman $\rho = -.390$, path coefficient = .32) suggest that more positive feelings connected with the timeliness, performance, and quality of the study obligations result in more positive satisfaction with online study and thus indirectly in lower levels of stress. These results were expected as they have been produced by previous, as well pre-COVID-19 studies [51,52].

The third hypothesis, that feelings toward online study (working conditions, pedagogy, and well-being) influence stress, was also confirmed. The results (Spearman $\rho = -.316$, path coefficient = .32) suggest that higher scores or less negative feelings connected with the timeliness, performance, and quality of the study obligations reduce stress. Similar results were found in a qualitative study conducted among university students during the COVID-19 lockdown, which showed that the quality of internet connections and the study environment were among the main sources of students' stress [61].

The results clearly show that more positive satisfaction with online study and more positive feelings toward study obligations during COVID-19 lockdowns are moderate predictors of stress. FETSOS is a negative predictor of SAT-5, meaning that positive feelings (opinions) toward the study obligations result in higher satisfaction and that lower satisfaction and more intense negative feelings toward the study obligations are predictors of higher stress. Academic struggles may therefore increase already elevated distress among the postsecondary population [62]. Based on the findings of this study, the levels of perceived stress are higher in female and bachelor's degree students. However, differences in the terms of effect sizes are small. The differences in FETSOS and SAT-5 in both categories are almost nonexistent in terms of effect size.

The results of this study show higher levels of perceived stress among Slovenian postsecondary students during the COVID-19 pandemic, although the fact that the survey was conducted during the exam period might also have influenced the stress levels. Moreover, an upward trend in feelings of stress was also reported before the pandemic. Data from the international research study "Mladina 2018–2019" [63] show a high increase in the share of Slovenian young people aged 14–29 years who reported, over a 5-year interval, feeling stressed most of the days of the week. Similarly, data from a national study titled "Health-Related Behavioural Style of Slovenian Residents" [64] show that approximately a quarter (23.2%) of Slovenian people aged between 25 and 74 years reported regular or daily feelings of stress, with stress more commonly reported by women and younger individuals (28.3% of those aged between 25 and 34 years). To obtain a better insight into the trend in perceived stress in postpandemic times, a longitudinal study that also includes other potential impact factors would be desirable.

Prolonged, recurrent stress and poor stress management are among the key factors in the deterioration of an individual's health, as they increase the risk of many diseases and disorders.

Conversely, reducing stressors and strengthening an individual's resilience to stress make an important contribution to maintaining and improving mental health. From the findings of this study, we can therefore recommend that implementing structures and measures targeted at stress reduction, such as the establishment of psychological support, is crucial, especially in cases where stress levels call for mental health treatment. However, any measures should target working conditions and pedagogy as well, which calls for the elimination of or at least a reduction in the obstacles identified by FETSOS. It is outside the scope of this paper to suggest practical measures to postsecondary educational institutions, but those related to pedagogy are completely within its domain. At the top of the scale is motivation. This should therefore be raised (or at least not decreased) by all means possible at institutional and individual levels, showing students that the future is not dark and without hope. There is also no excuse not to make expectations about outcomes clear, and to increase access to necessary resources, at least in digital format. Organizing an e-library, for example, should not present a big problem. Problems with internet connections and adequacy of the workspace are factors that were induced by the lockdown and confinement to the home environment. Stress would also be reduced by keeping campuses and dormitories open, and by providing an online structure and ensuring that libraries are close by. However, it can be reasonably expected that due to the higher virulence (even with lower severity) of the new strains of coronavirus [65], or even the possible emergence of new zoonotic viruses [66], the experience gained over the past two years should be incorporated into efforts to minimize the negative impacts of internet-based education.

Limitations

The main limitation of the study was the self-selection of the respondents. Even if a high number of respondents were included in the data set, they are representative of a population who responded to all of the items. It is therefore impossible to make inferences about the population of individuals who did not respond to the study. Furthermore, the fact that the study featured a preponderance of male respondents could have influenced the results, as certain mental health problems may be more frequent among female students. Although other mental health problems (anxiety, depression, etc) and other variables that interact with or influence stress are important, they were not included in any recent study and we did not report them in this paper. The missing links remain to be explored.

Directions for Future Research

We suggest that the survey be repeated after the lockdowns have ended to find out whether there are differences in levels of stress and the factors impacting it. Additionally, an international comparative study would shed light on the differences that are based on various sociocultural factors.

Conclusion

In conclusion, the side effects caused by FODE [3], as well as the ad hoc educational practices resulting from the pandemic [3], resulted in higher stress, which is known as 1 of the main causes of mental health problems, such as anxiety and

depression. We can only hope that these effects are just transitional. If they are not, society will see a higher number of people with mental health problems well into the future.

Acknowledgments

This publication arises from the “Measures in the Field of COVID-19 Spread Management with a Focus on Vulnerable Populations” project, funded by Slovenia and the European Social Fund, of the European Structural and Investment Funds, covered by Regulation (EU) No. 1303/2013 of the European Parliament and of the Council (the Common Provisions Regulation). The content of this paper represents the views of the authors only and is their sole responsibility; it cannot be considered to reflect the views of the European Commission or any other body of the European Union. The European Commission does not accept any responsibility for the use that may be made of the information it contains.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Questionnaire.

[[DOCX File , 35 KB - jmir_v24i3e30778_app1.docx](#)]

Multimedia Appendix 2

Measures of central tendencies (N=4455), communalities and factor loadings of the PSS-4 (N=2235). PSS-4: Perceived Stress Scale.

[[DOCX File , 14 KB - jmir_v24i3e30778_app2.docx](#)]

Multimedia Appendix 3

Measures of central tendencies (N=4455), communalities, and factor loadings of SAT-5 (N=2235). SAT-5: Satisfaction with Online Study Scale.

[[DOCX File , 14 KB - jmir_v24i3e30778_app3.docx](#)]

Multimedia Appendix 4

Measures of central tendencies (N=4455), communalities, and factor loadings of FETSOS (N=2235). FETSOS: Feelings Towards Study Obligations Scale.

[[DOCX File , 16 KB - jmir_v24i3e30778_app4.docx](#)]

References

1. Kumaraswamy N. Academic stress, anxiety and depression among college students: a brief review. *Int Rev Soc Soc Humanit* 2013;5(1):135-143 [[FREE Full text](#)]
2. Regehr C, Glancy D, Pitts A. Interventions to reduce stress in university students: a review and meta-analysis. *J Affect Disord* 2013 May 15;148(1):1-11. [doi: [10.1016/j.jad.2012.11.026](#)] [Medline: [23246209](#)]
3. Ploj-Virtič M, Dolenc K, Šorgo A. Changes in online distance learning behaviour of university students during the coronavirus disease 2019 outbreak, and development of the model of forced distance online learning preferences. *Eur J Educ Res* 2021 Jan 15;10(1):393-411. [doi: [10.12973/eu-jer.10.1.393](#)]
4. Dolenc K, Šorgo A, Ploj-Virtič M. The difference in views of educators and students on forced online distance education can lead to unintentional side effects. *Educ Inf Technol (Dordr)* 2021 Apr 28:1-27 [[FREE Full text](#)] [doi: [10.1007/s10639-021-10558-4](#)] [Medline: [33935580](#)]
5. Wathelet M, Duhem S, Vaiva G, Baubet T, Habran E, Veerapa E, et al. Factors associated with mental health disorders among university students in France confined during the COVID-19 pandemic. *JAMA Netw Open* 2020 Oct 01;3(10):e2025591 [[FREE Full text](#)] [doi: [10.1001/jamanetworkopen.2020.25591](#)] [Medline: [33095252](#)]
6. Odriozola-González P, Planchuelo-Gómez Á, Irurtia MJ, de Luis-García R. Psychological effects of the COVID-19 outbreak and lockdown among students and workers of a Spanish university. *Psychiatry Res* 2020 Aug;290:113108 [[FREE Full text](#)] [doi: [10.1016/j.psychres.2020.113108](#)] [Medline: [32450409](#)]
7. Rogowska AM, Pavlova I, Kuśnierz C, Ochnik D, Bodnar I, Petrytsa P. Does physical activity matter for the mental health of university students during the COVID-19 pandemic? *J Clin Med* 2020 Oct 29;9(11):3494 [[FREE Full text](#)] [doi: [10.3390/jcm9113494](#)] [Medline: [33138047](#)]
8. Jané-Llopis E, Anderson P, Segura L, Zabaleta E, Muñoz R, Ruiz G, et al. Mental ill-health during COVID-19 confinement. *BMC Psychiatry* 2021 Apr 14;21(1):194 [[FREE Full text](#)] [doi: [10.1186/s12888-021-03191-5](#)] [Medline: [33853562](#)]

9. Lawrence SA, Garcia J, Stewart C, Rodriguez C. The mental and behavioral health impact of COVID-19 stay at home orders on social work students. *Soc Work Educ* 2021 Feb 03;1-15. [doi: [10.1080/02615479.2021.1883582](https://doi.org/10.1080/02615479.2021.1883582)]
10. 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. [doi: [10.1016/S0140-6736\(20\)30309-3](https://doi.org/10.1016/S0140-6736(20)30309-3)] [Medline: [32043982](https://pubmed.ncbi.nlm.nih.gov/32043982/)]
11. DiGiovanni C, Conley J, Chiu D, Zaborski J. Factors influencing compliance with quarantine in Toronto during the 2003 SARS outbreak. *Biosecur Bioterror* 2004 Dec;2(4):265-272. [doi: [10.1089/bsp.2004.2.265](https://doi.org/10.1089/bsp.2004.2.265)] [Medline: [15650436](https://pubmed.ncbi.nlm.nih.gov/15650436/)]
12. Rajkumar RP. COVID-19 and mental health: a review of the existing literature. *Asian J Psychiatr* 2020 Aug;52:102066 [FREE Full text] [doi: [10.1016/j.ajp.2020.102066](https://doi.org/10.1016/j.ajp.2020.102066)] [Medline: [32302935](https://pubmed.ncbi.nlm.nih.gov/32302935/)]
13. Jewell JS, Farewell CV, Welton-Mitchell C, Lee-Winn A, Walls J, Leiferman JA. Mental health during the COVID-19 pandemic in the United States: online survey. *JMIR Form Res* 2020 Oct 23;4(10):e22043 [FREE Full text] [doi: [10.2196/22043](https://doi.org/10.2196/22043)] [Medline: [33006939](https://pubmed.ncbi.nlm.nih.gov/33006939/)]
14. Fu W, Yan S, Zong Q, Anderson-Luxford D, Song X, Lv Z, et al. Mental health of college students during the COVID-19 epidemic in China. *J Affect Disord* 2021 Feb 01;280(Pt A):7-10 [FREE Full text] [doi: [10.1016/j.jad.2020.11.032](https://doi.org/10.1016/j.jad.2020.11.032)] [Medline: [33197782](https://pubmed.ncbi.nlm.nih.gov/33197782/)]
15. Baloran ET. Knowledge, attitudes, anxiety, and coping strategies of students during COVID-19 pandemic. *J Loss Trauma* 2020 May 22;25(8):635-642. [doi: [10.1080/15325024.2020.1769300](https://doi.org/10.1080/15325024.2020.1769300)]
16. Marques G, Drissi N, Díez IT, de Abajo BS, Ouhbi S. Impact of COVID-19 on the psychological health of university students in Spain and their attitudes toward mobile mental health solutions. *Int J Med Inform* 2021 Mar;147:104369. [doi: [10.1016/j.ijmedinf.2020.104369](https://doi.org/10.1016/j.ijmedinf.2020.104369)] [Medline: [33388481](https://pubmed.ncbi.nlm.nih.gov/33388481/)]
17. Babicka-Wirkus A, Wirkus L, Stasiak K, Kozłowski P. University students' strategies of coping with stress during the coronavirus pandemic: data from Poland. *PLoS One* 2021 Jul 26;16(7):e0255041 [FREE Full text] [doi: [10.1371/journal.pone.0255041](https://doi.org/10.1371/journal.pone.0255041)] [Medline: [34310649](https://pubmed.ncbi.nlm.nih.gov/34310649/)]
18. Marcén-Román Y, Gasch-Gallen A, Vela Martín de la Mota II, Calatayud E, Gómez-Soria I, Rodríguez-Roca B. Stress perceived by university health sciences students, 1 year after COVID-19 pandemic. *Int J Environ Res Public Health* 2021 May 14;18(10):5233 [FREE Full text] [doi: [10.3390/ijerph18105233](https://doi.org/10.3390/ijerph18105233)] [Medline: [34069066](https://pubmed.ncbi.nlm.nih.gov/34069066/)]
19. Son C, Hegde S, Smith A, Wang X, Sasangohar F. Effects of COVID-19 on college students' mental health in the United States: interview survey study. *J Med Internet Res* 2020 Sep 03;22(9):e21279 [FREE Full text] [doi: [10.2196/21279](https://doi.org/10.2196/21279)] [Medline: [32805704](https://pubmed.ncbi.nlm.nih.gov/32805704/)]
20. Elmer T, Mepham K, Stadtfeld C. Students under lockdown: comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland. *PLoS One* 2020 Jul 23;15(7):e0236337 [FREE Full text] [doi: [10.1371/journal.pone.0236337](https://doi.org/10.1371/journal.pone.0236337)] [Medline: [32702065](https://pubmed.ncbi.nlm.nih.gov/32702065/)]
21. Yu Y, She R, Luo S, Xin M, Li L, Wang S, et al. Factors influencing depression and mental distress related to COVID-19 among university students in China: online cross-sectional mediation study. *JMIR Ment Health* 2021 Feb 22;8(2):e22705 [FREE Full text] [doi: [10.2196/22705](https://doi.org/10.2196/22705)] [Medline: [33616541](https://pubmed.ncbi.nlm.nih.gov/33616541/)]
22. Schröpfer K, Schmidt N, Kus S, Koob C, Coenen M. Psychological stress among students in health-related fields during the COVID-19 pandemic: results of a cross-sectional study at selected Munich universities. *Int J Environ Res Public Health* 2021 Jun 19;18(12):6611 [FREE Full text] [doi: [10.3390/ijerph18126611](https://doi.org/10.3390/ijerph18126611)] [Medline: [34205383](https://pubmed.ncbi.nlm.nih.gov/34205383/)]
23. Rogowska M, Kuśniercz C, Bokszczanin A. Examining anxiety, life satisfaction, general health, stress and coping styles during COVID-19 pandemic in Polish sample of university students. *PRBM* 2020 Sep;13:797-811 [FREE Full text] [doi: [10.2147/prbm.s266511](https://doi.org/10.2147/prbm.s266511)]
24. Gavurova B, Ivankova V, Rigelsky M. Relationships between perceived stress, depression and alcohol use disorders in university students during the COVID-19 pandemic: a socio-economic dimension. *Int J Environ Res Public Health* 2020 Nov 28;17(23):8853 [FREE Full text] [doi: [10.3390/ijerph17238853](https://doi.org/10.3390/ijerph17238853)] [Medline: [33260778](https://pubmed.ncbi.nlm.nih.gov/33260778/)]
25. Browning MHEM, Larson LR, Sharaievska I, Rigolon A, McAnirlin O, Mullenbach L, et al. Psychological impacts from COVID-19 among university students: risk factors across seven states in the United States. *PLoS One* 2021 Jan 7;16(1):e0245327 [FREE Full text] [doi: [10.1371/journal.pone.0245327](https://doi.org/10.1371/journal.pone.0245327)] [Medline: [33411812](https://pubmed.ncbi.nlm.nih.gov/33411812/)]
26. Kohls E, Baldofski S, Moeller R, Klemm S, Rummel-Kluge C. Mental health, social and emotional well-being, and perceived burdens of university students during COVID-19 pandemic lockdown in Germany. *Front Psychiatry* 2021 Apr 6;12:643957 [FREE Full text] [doi: [10.3389/fpsyg.2021.643957](https://doi.org/10.3389/fpsyg.2021.643957)] [Medline: [33889102](https://pubmed.ncbi.nlm.nih.gov/33889102/)]
27. Feng Y, Zhang Y. Evaluation and analysis of mental health level of college students with financial difficulties under the background of COVID-19. *Front Psychol* 2021 Apr 9;12:649195 [FREE Full text] [doi: [10.3389/fpsyg.2021.649195](https://doi.org/10.3389/fpsyg.2021.649195)] [Medline: [33897557](https://pubmed.ncbi.nlm.nih.gov/33897557/)]
28. Hodges C, Moore S, Lockee B, Trust T, Bond A. The Difference Between Emergency Remote Teaching and Online Learning. URL: <https://tinyurl.com/2p8d4hjh> [accessed 2022-03-09]
29. Khan A, Sultana M, Hossain S, Hasan M, Ahmed H, Sikder T. *J Affect Disord* 2020 Dec 01;277:121-128 [FREE Full text] [doi: [10.1016/j.jad.2020.07.135](https://doi.org/10.1016/j.jad.2020.07.135)] [Medline: [32818775](https://pubmed.ncbi.nlm.nih.gov/32818775/)]
30. Sternberg R. The theory of successful intelligence. *Rev Interam Psicol J Psychol* 2005:189-202 [FREE Full text] [doi: [10.1017/cbo9780511977244.026](https://doi.org/10.1017/cbo9780511977244.026)]

31. Organisation for Economic Co-operation and Development. Education at a Glance 2020. URL: https://www.oecd-ilibrary.org/education/education-at-a-glance-2020_69096873-en [accessed 2022-03-09]
32. Artino AR. Online or face-to-face learning? Exploring the personal factors that predict students' choice of instructional format. *Internet High Educ* 2010 Dec;13(4):272-276. [doi: [10.1016/j.iheduc.2010.07.005](https://doi.org/10.1016/j.iheduc.2010.07.005)]
33. Fletcher JD, Tobias S, Wisner RA. Learning anytime, anywhere: advanced distributed learning and the changing face of education. *Educ Res* 2016 Nov 30;36(2):96-102. [doi: [10.3102/0013189x07300034](https://doi.org/10.3102/0013189x07300034)]
34. Gabrovec B, Selak Š, Crnkovič N, Cesar K, Šorgo A. Raziskava o doživljanju epidemije COVID-19 med študenti: poročilo o opravljeni raziskavi. URL: https://www.nijz.si/sites/www.nijz.si/files/uploaded/splosno_porocilo_o_opravljeni_raziskavi_-_raziskava_o_dozivljanju_epidemije_covid-19_med_studenti.pdf [accessed 2022-03-09]
35. 1KA OneClick survey. URL: <https://www.1ka.si/d/en> [accessed 2022-01-17]
36. Jeriček KH, Roškar M, Drev A. Z zdravjem povezana vedenja v šolskem obdobju med mladostniki v Sloveniji, izsledki mednarodne raziskave HBSC, 2018. URL: <https://www.nijz.si/sl/publikacije/z-zdravjem-povezana-vedenja-v-solskem-obdobju-med-mladostniki-v-sloveniji-izsledki> [accessed 2022-03-09]
37. Cohen S. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. *The Social Psychology of Health*. Thousand Oaks, CA: SAGE Publications; 1988:31-67.
38. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. *J Health Soc Behav* 1983 Dec;24(4):385. [doi: [10.2307/2136404](https://doi.org/10.2307/2136404)]
39. Kavčič T, Avsec A, Zager Kocjan G. Psychological functioning of Slovene adults during the COVID-19 pandemic: does resilience matter? *Psychiatr Q* 2021 Mar 17;92(1):207-216 [FREE Full text] [doi: [10.1007/s11126-020-09789-4](https://doi.org/10.1007/s11126-020-09789-4)] [Medline: [32556914](https://pubmed.ncbi.nlm.nih.gov/32556914/)]
40. Debevc M, Weiss J, Šorgo A, Kožuh I. Solfeggio learning and the influence of a mobile application based on visual, auditory and tactile modalities. *Br J Educ Technol* 2019 Apr 18;51(1):177-193. [doi: [10.1111/bjet.12792](https://doi.org/10.1111/bjet.12792)]
41. Beck LA, Csikszentmihalyi M. Flow: the psychology of optimal experience. *J Leis Res* 2018 Feb 13;24(1):93-94. [doi: [10.1080/00222216.1992.11969876](https://doi.org/10.1080/00222216.1992.11969876)]
42. Nijs L, Leman M. Interactive technologies in the instrumental music classroom: a longitudinal study with the Music Paint Machine. *Comput Educ* 2014 Apr;73:40-59. [doi: [10.1016/j.compedu.2013.11.008](https://doi.org/10.1016/j.compedu.2013.11.008)]
43. Patil VH, Singh SN, Mishra S, Donovan DT. Parallel Analysis Engine to Aid in Determining Number of Factors to Retain using R [Computer software]. URL: <https://analytics.gonzaga.edu/parallelengine/> [accessed 2022-03-09]
44. Kline RB. In: Kenny DA, Little DT, editors. *Principles and Practice of Structural Equation Modeling: Fourth Edition*. New York, NY: Guilford Press; Nov 2015:509-512.
45. Byrne BM. *Structural Equation Modeling With AMOS: Basic Concepts, Applications, and Programming, Third Edition*. Oxfordshire, UK: Routledge; 2016.
46. Gaskin J, Lim J. "Model Fit Measures", AMOS Plugin. 2016. URL: <http://statwiki.gaskin.com/index.php?title=Plugins> [accessed 2021-05-20]
47. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct Equ Modeling* 1999 Jan;6(1):1-55. [doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)]
48. Mondo M, Sechi C, Cabras C. Psychometric evaluation of three versions of the Italian Perceived Stress Scale. *Curr Psychol* 2019 Jan 08;40(4):1884-1892. [doi: [10.1007/s12144-019-0132-8](https://doi.org/10.1007/s12144-019-0132-8)]
49. Prowse R, Sherratt F, Abizaid A, Gabrys RL, Hellemans KGC, Patterson ZR, et al. Coping with the COVID-19 pandemic: examining gender differences in stress and mental health among university students. *Front Psychiatry* 2021 Apr 7;12:650759 [FREE Full text] [doi: [10.3389/fpsy.2021.650759](https://doi.org/10.3389/fpsy.2021.650759)] [Medline: [33897499](https://pubmed.ncbi.nlm.nih.gov/33897499/)]
50. Arnett J. Emerging adulthood: a theory of development from the late teens through the twenties. *Am Psychol* 2000;55(5):469-480 [FREE Full text] [doi: [10.1037/0003-066x.55.5.469](https://doi.org/10.1037/0003-066x.55.5.469)]
51. Zhai Y, Du X. Addressing collegiate mental health amid COVID-19 pandemic. *Psychiatry Res* 2020 Jun;288:113003 [FREE Full text] [doi: [10.1016/j.psychres.2020.113003](https://doi.org/10.1016/j.psychres.2020.113003)] [Medline: [32315885](https://pubmed.ncbi.nlm.nih.gov/32315885/)]
52. Song L, Singleton E, Hill J, Koh M. Improving online learning: student perceptions of useful and challenging characteristics. *Internet High Educ* 2004 Jan;7(1):59-70 [FREE Full text] [doi: [10.1016/j.iheduc.2003.11.003](https://doi.org/10.1016/j.iheduc.2003.11.003)]
53. Husky MM, Kovess-Masfety V, Swendsen JD. Stress and anxiety among university students in France during Covid-19 mandatory confinement. *Compr Psychiatry* 2020 Oct;102:152191 [FREE Full text] [doi: [10.1016/j.comppsy.2020.152191](https://doi.org/10.1016/j.comppsy.2020.152191)] [Medline: [32688023](https://pubmed.ncbi.nlm.nih.gov/32688023/)]
54. Hakami Z, Vishwanathaiah S, Abuzinadah SH, Alhaddad AJ, Bokhari AM, Marghalani HY, et al. Effects of COVID-19 lockdown on the mental health of dental students: a longitudinal study. *J Dent Educ* 2021 Aug 13:1-14 [FREE Full text] [doi: [10.1002/jdd.12758](https://doi.org/10.1002/jdd.12758)] [Medline: [34387880](https://pubmed.ncbi.nlm.nih.gov/34387880/)]
55. Alvi T, Assad F, Ramzan M, Khan F. Depression, anxiety and their associated factors among medical students. *J Coll Physicians Surg Pak* 2010 Feb;20(2):122-126. [Medline: [20378041](https://pubmed.ncbi.nlm.nih.gov/20378041/)]
56. Gao W, Ping S, Liu X. Gender differences in depression, anxiety, and stress among college students: a longitudinal study from China. *J Affect Disord* 2020 Feb 15;263:292-300. [doi: [10.1016/j.jad.2019.11.121](https://doi.org/10.1016/j.jad.2019.11.121)] [Medline: [31818792](https://pubmed.ncbi.nlm.nih.gov/31818792/)]

57. Pieh C, Budimir S, Probst T. The effect of age, gender, income, work, and physical activity on mental health during coronavirus disease (COVID-19) lockdown in Austria. *J Psychosom Res* 2020 Dec;136:110186 [FREE Full text] [doi: [10.1016/j.jpsychores.2020.110278](https://doi.org/10.1016/j.jpsychores.2020.110278)] [Medline: [33075602](https://pubmed.ncbi.nlm.nih.gov/33075602/)]
58. Besser A, Flett GL, Zeigler-Hill V. Adaptability to a sudden transition to online learning during the COVID-19 pandemic: understanding the challenges for students. *Scholarsh Teach Learn Psychol* 2020 Oct 19. [doi: [10.1037/stl0000198](https://doi.org/10.1037/stl0000198)]
59. Aucejo EM, French J, Ugalde Araya MP, Zafar B. The impact of COVID-19 on student experiences and expectations: evidence from a survey. *J Public Econ* 2020 Nov;191:104271 [FREE Full text] [doi: [10.1016/j.jpubeco.2020.104271](https://doi.org/10.1016/j.jpubeco.2020.104271)] [Medline: [32873994](https://pubmed.ncbi.nlm.nih.gov/32873994/)]
60. Lee J, Jang S. An exploration of stress and satisfaction in college students. *Serv Mark Q* 2015 Jul 17;36(3):245-260. [doi: [10.1080/15332969.2015.1046774](https://doi.org/10.1080/15332969.2015.1046774)]
61. Al-Kumaim NH, Alhazmi AK, Mohammed F, Gazem NA, Shabbir MS, Fazea Y. Exploring the impact of the COVID-19 pandemic on university students' learning life: an integrated conceptual motivational model for sustainable and healthy online learning. *Sustainability* 2021 Feb 26;13(5):2546. [doi: [10.3390/su13052546](https://doi.org/10.3390/su13052546)]
62. Kecojevic A, Basch CH, Sullivan M, Davi NK. The impact of the COVID-19 epidemic on mental health of undergraduate students in New Jersey, cross-sectional study. *PLoS One* 2020;15(9):e0239696 [FREE Full text] [doi: [10.1371/journal.pone.0239696](https://doi.org/10.1371/journal.pone.0239696)] [Medline: [32997683](https://pubmed.ncbi.nlm.nih.gov/32997683/)]
63. Naterer A, Lavrič M, Klanjšek R, Flere S, Rutar T, Lahe D, et al. Slovenska Mladina 2018-2019. URL: https://mlad.si/uploads/objave/NOVICE_UREDNIŠTVO/Slovenska_mladina_2018-2019.pdf.pdf [accessed 2022-03-09]
64. Vinko M, Kofol BT, Koročec A, Tomšič S, Vrdelja M. Kako Skrbimo Za Zdravje? Z zdravjem povezan vedenjski slog prebivalcev Slovenije 2016. URL: https://www.nijz.si/sites/www.nijz.si/files/publikacije-datoteke/kako_skrbimo_za_zdravje_splet_3007_koncna.pdf [accessed 2022-03-09]
65. Fokas AS, Kastis GA. SARS-CoV-2: the second wave in Europe. *J Med Internet Res* 2021 May 18;23(5):e22431 [FREE Full text] [doi: [10.2196/22431](https://doi.org/10.2196/22431)] [Medline: [33939621](https://pubmed.ncbi.nlm.nih.gov/33939621/)]
66. Benvenuto D, Giovanetti M, Ciccozzi A, Spoto S, Angeletti S, Ciccozzi M. The 2019-new coronavirus epidemic: evidence for virus evolution. *J Med Virol* 2020 Apr 07;92(4):455-459 [FREE Full text] [doi: [10.1002/jmv.25688](https://doi.org/10.1002/jmv.25688)] [Medline: [31994738](https://pubmed.ncbi.nlm.nih.gov/31994738/)]

Abbreviations

- CFA:** confirmatory factor analysis
- EFA:** exploratory factor analysis
- FETSOS:** Feelings Towards Study Obligations Scale
- FODE:** forced online distance education
- PAF:** principal axis factoring
- PSS-4:** Perceived Stress Scale
- SAT-5:** Satisfaction with Online Study Scale
- SEM:** structural equation modeling

Edited by A Mavragani; submitted 28.05.21; peer-reviewed by RP Rajkumar, S Gordon; comments to author 30.07.21; revised version received 23.09.21; accepted 05.02.22; published 15.03.22.

Please cite as:

Šorgo A, Crnkovič N, Gabrovec B, Cesar K, Selak Š

Influence of Forced Online Distance Education During the COVID-19 Pandemic on the Perceived Stress of Postsecondary Students: Cross-sectional Study

J Med Internet Res 2022;24(3):e30778

URL: <https://www.jmir.org/2022/3/e30778>

doi: [10.2196/30778](https://doi.org/10.2196/30778)

PMID: [35171098](https://pubmed.ncbi.nlm.nih.gov/35171098/)

©Andrej Šorgo, Nuša Crnkovič, Branko Gabrovec, Katarina Cesar, Špela Selak. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org/>), 15.03.2022. 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.

Original Paper

The Impact of COVID-19 Lockdown on Daily Activities, Cognitions, and Stress in a Lonely and Distressed Population: Temporal Dynamic Network Analysis

Matthias Haucke^{1,2}, BSc, MSc; Andreas Heinz¹, MD, PhD; Shuyan Liu^{1*}, BSc, MSc, PhD; Stephan Heinzel^{2*}, Dipl-Psych, PhD

¹Department of Psychiatry and Neurosciences, Charité–Universitätsmedizin Berlin, Campus Charité Mitte, Berlin, Germany

²Clinical Psychology and Psychotherapy, Department of Education and Psychology, Freie Universität Berlin, Berlin, Germany

*these authors contributed equally

Corresponding Author:

Matthias Haucke, BSc, MSc

Department of Psychiatry and Neurosciences

Charité–Universitätsmedizin Berlin

Campus Charité Mitte

Charitéplatz 1

Berlin, 10117

Germany

Phone: 49 015155710318

Email: matthias.haucke@fu-berlin.de

Abstract

Background: The COVID-19 pandemic and its associated lockdown measures impacted mental health worldwide. However, the temporal dynamics of causal factors that modulate mental health during lockdown are not well understood.

Objective: We aimed to understand how a COVID-19 lockdown changes the temporal dynamics of loneliness and other factors affecting mental health. This is the first study that compares network characteristics between lockdown stages to prioritize mental health intervention targets.

Methods: We combined ecological momentary assessments with wrist-worn motion tracking to investigate the mechanism and changes in network centrality of symptoms and behaviors before and during lockdown. A total of 258 participants who reported at least mild loneliness and distress were assessed 8 times a day for 7 consecutive days over a 213-day period from August 8, 2020, through March 9, 2021, in Germany, covering a “no-lockdown” and a “lockdown” stage. COVID-19–related worry, information-seeking, perceived restriction, and loneliness were assessed by digital visual analog scales ranging from 0 to 100. Social activity was assessed on a 7-point Likert scale, while physical activity was recorded from wrist-worn actigraphy devices.

Results: We built a multilevel vector autoregressive model to estimate dynamic networks. To compare network characteristics between a no-lockdown stage and a lockdown stage, we performed permutation tests. During lockdown, loneliness had the highest impact within the network, as indicated by its centrality index (ie, an index to identify variables that have a strong influence on the other variables). Moreover, during lockdown, the centrality of loneliness significantly increased. Physical activity contributed to a decrease in loneliness amid the lockdown stage.

Conclusions: The COVID-19 lockdown increased the central role of loneliness in triggering stress-related behaviors and cognition. Our study indicates that loneliness should be prioritized in mental health interventions during lockdown. Moreover, physical activity can serve as a buffer for loneliness amid social restrictions.

(*J Med Internet Res* 2022;24(3):e32598) doi:[10.2196/32598](https://doi.org/10.2196/32598)

KEYWORDS

COVID-19; mental health; outbreak; epidemic; pandemic; psychological response; emotional well-being; ecological momentary assessment; risk; protective factors; lockdown measures; loneliness; mood inertia; stressors; mobile apps; mHealth; digital health; EMA; smartphone apps; network model; cognition; stress; temporal dynamic network; permutation testing; network comparison; network characteristics; multilevel vector autoregressive model; mlVAR

Introduction

The outbreak of COVID-19 is an unprecedented global health challenge; as of November 2021, there are 259,502,031 confirmed cases and 5,183,003 deaths globally [1]. To mitigate the spread of SARS-CoV-2, most countries enforced lockdown measures, including social restrictions, travel bans, stay-at-home orders, and business shutdown. Together with the pandemic per se, these lockdown measures increased global mental health problems [2,3]. Reasons for this are an increase of distress or loneliness during the COVID-19 lockdown [4-7], yet most studies are overlooking the directionality between behavior and cognition over time. Recently, a network approach to psychopathology proposed that changes in mental health result from a temporal dynamic interaction between mental states, such that one mental state at one moment in time (eg, worry) can trigger other mental states at the next moment in time (eg, feeling stressed) [8]. We set out to examine whether lockdown measures can alter the dynamic network structure of behavior (eg, physical activity) and pandemic-related mental states (eg, worry). To do so, we compared differences between moment-to-moment time-lagged associations of pandemic-related cognitions, behaviors, and mental health, and tested for changes in centrality between lockdown stages. Comparing centrality (ie, an index to identify variables that have a strong influence on the other variables) can be informative in finding the most protective or detrimental temporal influence on mental health amid a lockdown [9,10]. This knowledge can be transferred to prioritize targets for pandemic-related mental health care interventions.

Psychological distress and social isolation are risk factors for developing mental disorders [11-15]. Therefore, we focused on a subpopulation who were experiencing at least mild levels of psychological distress and loneliness amid the COVID-19 pandemic. Moreover, we gathered real-life data using ecological momentary assessments (EMAs) via smartphone technology and measured objective physical activity via wrist-worn actigraphy devices. We investigated the temporal associations between loneliness, stress, physical and social activity, and COVID-19-related behaviors and cognitions.

We measured three COVID-19-related cognitions: perceived restriction in everyday life due to the pandemic, seeking information about the pandemic, and worrying about the pandemic's impact on one's life. Worries about the COVID-19-related economic downfall and the possible health impact on oneself or others can increase psychological distress [7,16]. In addition, distress, anxiety, depression, and anger are further increased by physical and social distancing measures [17,18]. People who stayed at home often acquired more COVID-19-related information through digital media, which increased anxiety and psychological distress [19-22]. Thus, COVID-19-related worrying, perceptions of restrictions, and information-seeking can be central causes of mental health issues.

Prior to the COVID-19 pandemic, loneliness was already recognized as one of the most pressing issues in modern societies [23]. Loneliness is an aversive state resulting from a

discrepancy between an individual's desired and realized social relationships [24]. Limiting social contacts and closing off social spaces can help to halt the spread of COVID-19; however, they also increase feelings of loneliness [7,25]. Loneliness has serious consequences for health, including increasing the risk of cardiovascular disease and immune dysfunction, depression, anxiety, and suicidal ideation [26]. To buffer against feelings of loneliness during lockdown, it can be essential to receive social support and engage in digital social activities [27,28].

A second buffer against mental health problems during the pandemic might be physical activity. Physical activity can relieve stress [29]; enhance cognitive abilities [30]; and reduce the risk of diabetes [31], cardiovascular disease [32], cancer [33], and mental disorders [34,35]. Conversely, sedentary behavior, defined as low-energy-expenditure behavior (≤ 1.5 metabolic equivalents), increases the risk for negative health outcomes, including type 2 diabetes mellitus, cardiovascular disease, and all-cause mortality [36-38]. Physical activity can lead to physiological reactions associated with decreased depression, such as an increase in neuroplasticity, cerebral blood flow, delivery of neurotrophic factors and oxygen, and resistance to oxidative stress [39]. Finally, exercise can improve self-efficacy and self-esteem [40]. We assessed physical activity through actigraphy (ie, a wrist-worn device that obtains objective measures of physical activity) [41].

Our study was performed in Germany during a no-lockdown stage (August 8 to November 1, 2020) and a lockdown stage (November 2, 2020, to March 9, 2021). During the no-lockdown stage, the restrictions were lenient (eg, no private or public meeting restrictions, and leisure facilities, bars, and catering facilities were open). To counter the steep increase in active COVID-19 cases, the German government announced a lockdown on November 2, 2020, including social restrictions, travel bans, closing of restaurants and cinemas, and business shutdowns. In addition, these lockdown measures were further tightened on December 16 (eg, closing of most retail; see Supplement A in [Multimedia Appendix 1](#)).

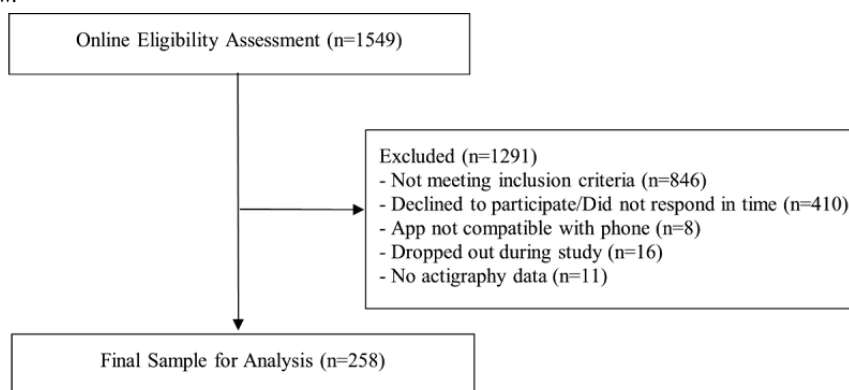
The aim of this study was to examine the temporal dynamic interplay between COVID-19 pandemic-related cognitions, behaviors, and mental health states. This is the first study to use a dynamic network approach to compare moment-to-moment time-lagged associations between pandemic-related cognitions, behaviors, and mental health states between lockdown stages. Moreover, we examined whether the lockdown affects the centrality of loneliness and specific pandemic-related behaviors and cognitions (ie, a more central variable has more and stronger connections to other variables). This helps to identify the most protective or detrimental influences on mental health during a lockdown. This knowledge, in turn, can be used to prioritize mental health intervention targets. Specifically, we hypothesized that a lockdown, in comparison to a no-lockdown period, increases the centrality of stress, physical activity, social contacts, and loneliness. Finally, we hypothesized that stress and loneliness will have a stronger influence on COVID-19-related behaviors and cognitions during lockdown than during no-lockdown.

Methods

Participants and Sampling

We assessed 1549 participants for eligibility in an online questionnaire. The final sample size was 258 (see Figure 1 for the recruitment flow). On average, participants missed 17.5%

Figure 1. Recruitment flow.



Inclusion criteria were (1) a minimum age of 18 years, (2) not working a night shift, (3) not being infected by COVID-19, (4) using an Android smartphone, and (5) speaking fluent German. Moreover, we targeted individuals who reported (6) perceived mild to moderate psychological distress and (7) sometimes felt lonely during the COVID-19 pandemic. We used the COVID-19 Peritraumatic Distress Index (CPDI [43]; cut-off score=28, indicating mild distress) questionnaire and the short-form of the University of California Los Angeles Loneliness Scale (ULS-8 [44]; cut-off score=16, indicating mild loneliness), respectively. The CPDI was designed for evaluating changes in mental health status, cognitive skills, avoidance and compulsive behavior, physical symptoms, and loss of social functioning due to the COVID-19 pandemic. The questionnaire has been previously validated in Germany [43].

Study Design and Procedure

The study was conducted in Germany over a 213-day period between August 8, 2020, and March 9, 2021, covering a no-lockdown and a lockdown stage. Participants were recruited via online advertisements on university websites, Twitter, and eBay classifieds. Participants had to fill in an online screening questionnaire on the Siuvo Intelligent Psychological Assessment Platform. After an initial contact via phone or email, we sent participants our study information, accelerometer, informed consent, and a QR code (to install a smartphone app) by mail. After they completed the study, participants sent back the study material by mail.

We conducted the EMA via the smartphone app “movisensXS” (movisens GmbH, Karlsruhe, Germany) developed for research purposes. This app is compliant with the General Data Protection Regulation (European Union) and Berlin Data Protection Act (Berliner Datenschutzgesetz). The app consists of a sociodemographic assessment (eg, age, gender, and years of education) and measures participants’ current experiences in real time. Participants filled in questionnaires for 7 consecutive days, in which they received 8 prompts (randomized within 1 hour and 45-minute blocks between 8 AM and 10 PM). We

of the questionnaires, no participants missed more than 50% of the sent questionnaires, and 117 data points were marked by the GGIR package [42] as “nonwear” and subsequently excluded from the analyses. Specifically, the accelerometer nonwear score was estimated based on the standard deviation and range of the raw data from each accelerometer axis [42].

performed an EMA that involves repeated sampling of individuals’ current behaviors and experiences in real time and in their natural environments. EMA minimizes recall bias, maximizes ecological validity, and allows approximating temporal causality (ie, Granger causality) [45]. A time series X is said to Granger-cause Y if it can be shown, usually through a series of t tests and F tests on lagged values of X (and with lagged values of Y also included), that the X values provide statistically significant information about future values of Y [46].

Moreover, we measured physical activity via the “GENEActiv” Original (Activinsights) monitor (dynamic range ± 8 g, sampling frequency range 10-100 Hz). Participants wore the actigraphy device on the left wrist.

Ethical Considerations

The study was approved by the ethics committee at Charité–Universitätsmedizin Berlin (reference EA2/143/20) and Freie Universität Berlin (reference 030/2020).

Measures

EMA Items

Stress was measured with the following question: “In this moment I feel stressed.” Other items started with “During the last hour...” followed by “to which extent did you feel constrained by the pandemic in your everyday life?” (perceived restriction), “to which extent did you worry about how the pandemic affects your personal situation?” (worry), “to which extent did you seek information about the Corona pandemic?” (information-seeking), and “to which extent did you feel lonely” (loneliness). Each of these items was measured on a visual analog scale (0-100: 0=not at all, 100=most frequent or severe). Duration of social activity was measured with the question “How long did your last social contact last?” via a Likert scale ranging from 1=“0 minutes” to 7=“50-60 minutes.”

Actigraphy Data

Physical activity data were collected using the actigraphy devices worn by each participant on the left wrist.

Statistical Analysis

Overview

All analyses were performed using R statistical software (version 3.5.3). In this section, we describe the data preparation procedures, averaged values of our measured items, estimation of the dynamic networks, and the permutation procedure used to test for group differences in centrality indices and dynamic association.

Data Preparation

We calculated the Euclidean norm (vector magnitude) of the raw signals of the three-measurement axis, which is a summary score of body acceleration and a validated measure for physical activity [47]. The Euclidean norm minus one (ENMO) is defined as $r_i - 1000$ [48], where



The actigraphy data from GENEActiv (100 Hz; .bin files) were downloaded using GENEActiv PC software V3.3. The GENEActiv .bin files were then exported into R statistical software V4.0.3 for processing using the GGIR package V1.2-0. We autocalibrated the raw triaxial accelerometer signals and computed the average ENMO metric for 1 hour before each beep. To exclude time frames in which participants did not wear their actigraphy device, we used the nonwear score of the GGIR package. We excluded time frames above the cut-off score of 1. As the EMA items were nonnormally distributed, we transformed all variables using the nonparanormal transformation [49]. To test for nonstationarity, we calculated a two-level autoregressive model for each lockdown group, in which each score of the variable included in our model was regressed on the immediately preceding score of that variable (ie, moment-to-moment inertia). A moment-to-moment inertia value larger than 1 indicates a nonstationary process [50]. We assumed stationarity, as the average moment-to-moment inertia ranged between 0.13 and 0.37 for all 7 included variables for

each lockdown group (see Supplement B in [Multimedia Appendix 1](#)). In addition, a Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was performed separately for every subject and variable. The KPSS test indicated that the data were stationary (approximately 99.9%). The R code of the statistical analyses is available online [51].

Dynamic Network Estimation

We built a first-order vector autoregressive model (VAR) with the R package mlVAR. Each variable at time point t was predicted by all variables (including itself) at the next time point of measurement (lag 1). The results of the network models consisted of nodes (variables) and directed edges (statistical relations) that were visualized via the R package qgraph [52]

Permutation Testing of Centrality Indices and Edge Differences

Permutation tests were used to compare individual path and network centrality between the lockdown and no-lockdown stages. The permutation procedure was developed by Wolfgang Viechtbauer and compares the results of the observed data with a distribution derived from repeated permutation (100,000) of the data under the null hypothesis [53-55]. To assess the importance of specific variables in the network of two groups, in-strength and out-strength were calculated from all (including nonsignificant) edges in the network. In-strength reflects the sum of ingoing edge weights, whereas out-strength reflects the sum of outgoing edge weights to the specific node [56,57]. A detailed description of the permutation procedures can be found in Supplement C in [Multimedia Appendix 1](#).

Results

Sociodemographics

Sociodemographic characteristics of the final sample (N=258), as well as results of independent t tests or χ^2 tests comparing these characteristics between a no-lockdown and lockdown stage are shown in [Table 1](#). As we had more women in our lockdown group, we tested the effect of gender on all measured variables. We found that, except for social duration, gender did not significantly affect our variables (see Supplement G in [Multimedia Appendix 1](#)).

Table 1. Sociodemographic characteristics of participants.

Characteristic	Total (August 8, 2020, to March 9, 2021; N=258)	No-lockdown period (August 8 to November 1, 2020; n=131)	Lockdown period (November 2 to March 9, 2021; n=127)	P value ^a
Age (years), mean (SD)	30.78 (11.16)	31.18 (10.52)	30.16 (11.67)	.55
Education (years), mean (SD)	15.28 (3.69)	15.1 (3.69)	15.46 (3.69)	.44
Gender, n (%)				.008
Male	77 (29.8)	49 (37.4)	28 (22.0)	
Female	178 (70.0)	82 (62.6)	96 (75.6)	
Diverse	3 (1.2)	0 (0)	3 (2.4)	
Family status, n (%)				.93
Single	114 (44.2)	61 (46.6)	53 (41.7)	
In relationship	92 (35.7)	45 (34.4)	47 (37.0)	
Married	48 (18.6)	23 (17.6)	25 (19.7)	
Other	4 (1.6)	2 (1.5)	2 (1.6)	
Number of children, mean (SD)	1.77 (0.78)	1.7 (0.78)	1.88 (0.78)	.38
Number living with others, mean (SD)	2.56 (2.15)	2.5 (1.29)	2.62 (2.77)	.65
Health status (1=very bad, 5=very good), mean (SD)	3.74 (0.86)	3.65 (0.91)	3.83 (0.81)	.09
COVID-19 risk group, n (%)	64 (24.8)	33 (25.2)	31 (24.4)	.80
COVID-19 distress (CPDI ^b), mean (SD)	47.56 (14.79)	48.32 (16.34)	46.76 (13.31)	.41
Loneliness (ULS-8 ^c), mean (SD)	22.57 (3.97)	22.01 (4.01)	23.15 (3.85)	.02

^aBased on independent *t* test or χ^2 test; unequal variance was assumed, and we applied the Welsh approximation to the degrees of freedom.

^bCPDI: COVID-19 Peritraumatic Distress Index.

^cULS-8: University of California Los Angeles Loneliness Scale.

Average-Based Lockdown Differences

To compare the no-lockdown and lockdown stages, we performed independent *t* tests using overall averages for each person. As shown in Table 2, the lockdown significantly

increased COVID-19 worries, perceived restriction, and duration of social contacts. Moreover, the lockdown significantly decreased physical activity. There was no statistically significant influence of lockdown on information-seeking, stress, and loneliness.

Table 2. Differences between no-lockdown and lockdown stages.

Variables	No-lockdown period (n=131), mean (SD)	Lockdown period (n=127), mean (SD)	P value ^a
EMAb items			
Loneliness	22.62 (20.82)	21.45 (19.80)	.64
COVID-19 worries	24.59 (18.36)	29.12 (17.33)	.04
COVID-19 perceived restriction	23.86 (17.83)	28.16 (17.05)	.05
COVID-19 information-seeking	22.85 (15.57)	23.46 (13.94)	.74
Social contacts	2.64 (0.95)	3.05 (1.00)	<.001
Stress	35.05 (18.43)	33.25 (17.34)	.42
Physical activity from actigraphy (microgravity)	40.15 (13.37)	35.24 (11.42)	.002

^a*t* test; unequal variance was assumed and we applied the Welsh approximation to the degrees of freedom.

^bEMA: ecological momentary assessment.

Network Estimation

We wanted to investigate how a lockdown affects the temporal dynamics of pandemic-related cognitions, behaviors, and mental

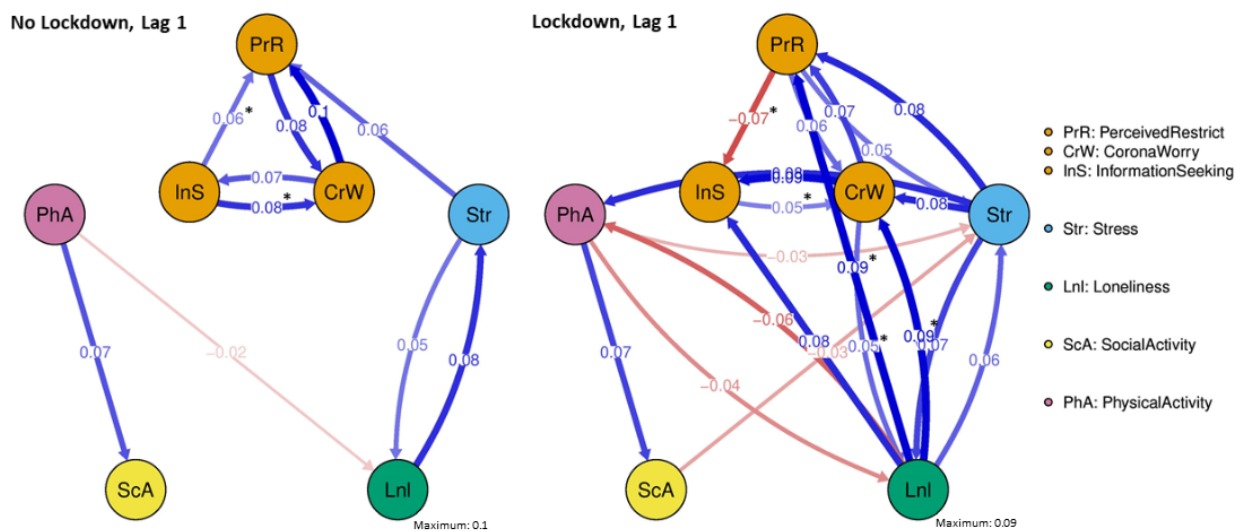
health states. To do so, we first estimated the temporal (ie, time-lagged) and bidirectional associations between detrimental and beneficial factors via multilevel VAR models [58-60]. These VAR models were then used to estimate temporal dynamic

networks for a lockdown and a no-lockdown stage. Permutation testing was used to test for group differences in individual paths and the network centrality of pandemic-related detrimental and beneficial mental health factors between the lockdown and no-lockdown stage. Moreover, the exploratory results of a permutation test for the difference in overall connectivity are provided in Supplement C of [Multimedia Appendix 1](#).

Edge Differences Between Groups

Figure 2 displays the “full” dynamic symptom networks for the lockdown and no-lockdown groups, which include only statistically significant *edges* (ie, time-lagged partial correlations with $\alpha < .05$). Permutation tests revealed that 7 of the edges were significantly different between the no-lockdown and lockdown groups at the uncorrected α level (indicated with an asterisk in **Figure 2**).

Figure 2. Temporal dynamic networks for a no-lockdown and a lockdown stage. Temporal relations among ecological momentary assessment and physical activity data, measured by actigraphy devices, estimated with a multilevel vector autoregressive model, and depicted as a graph where nodes are variables and edges (arrows connecting nodes) are statistically significant ($\alpha < .05$) partial correlations among variables. Thicker and more saturated edges depict stronger relations; positive relations are in blue and negative relations are in red. Associations that are significantly different between the no-lockdown and lockdown stages (permutation testing using a two-sided *P* value at the uncorrected α level) are marked with an asterisk.



Compared to no-lockdown, in a lockdown, participants showed a stronger connection from “loneliness” to “perceived restriction” (difference -0.114 , $P < .001$) and from “loneliness” to “COVID-19–related worry” (difference -0.0767 , $P = .03$).

Compared to no-lockdown, in a lockdown, participants showed a weaker connection from “information-seeking” to “perceived restriction” (difference 0.0609 , $P = .02$) and from “information-seeking” to “COVID-19–related worry” (difference 0.0477 , $P = .05$). In addition, information-seeking led to less information-seeking in the next moment (ie, weaker autocorrelation; difference 0.0754 , $P = .02$).

Compared to no-lockdown, during the lockdown, participants showed a stronger connection from “COVID-19–related worry” to “loneliness” (difference -0.0444 , $P = .05$).

Compared to no-lockdown, during the lockdown, participants showed a weaker connection from “perceived restriction” to “social activity” (difference 0.0065 , $P = .01$).

More information on the time-lagged partial correlations (ie, edges) that were significantly different during the lockdown can be found in [Table 3](#) (all, including nonsignificant, edge differences are shown in Supplement F of [Multimedia Appendix 1](#)).

Table 3. Significant edge differences of time-lagged partial correlation coefficients between the lockdown and no-lockdown stages.

Predictor (1-lag)	Outcome	Partial correlation coefficient		Difference in partial correlation coefficient	P value
		No-Lockdown	Lockdown		
Information-seeking	Perceived restriction	0.0548	-0.0062	0.0609	.02
Loneliness	Perceived restriction	0.001	0.115	-0.114	<.001
Information-seeking	COVID-19–related worry	0.0689	0.0212	0.0477	.05
Loneliness	COVID-19–related worry	0.0274	0.1042	-0.0767	.03
Information-seeking	Information-seeking	0.1721	0.0967	0.0754	.02
COVID-19–related worry	Loneliness	-0.0129	0.0315	-0.0444	.05
Perceived restriction	Social activity	0.0043	-0.0021	0.0065	.01

Centrality Indices

In-strength is the sum of *ingoing* edge weights to a specific node and *out-strength* is the sum of the *outgoing* edge weights to a specific node. During the no-lockdown stage, worrying about COVID-19 had the highest out-strength, indicating that when a participant reports worries about COVID-19 at one measurement occasion, it is likely that this participant will report other COVID-19–related behaviors and cognitions at the next measurement occasion. During lockdown, loneliness had the highest out-strength, indicating that when a participant reports feeling lonely in one moment, this participant is likely to report

COVID-19–related behaviors and cognitions in the next momentary assessment.

Permutation tests revealed a significant higher out-strength for “loneliness” during lockdown (difference -0.1975 , $P=.04$) and significant lower out-strength for “information-seeking” (difference 0.1452 , $P=.03$) at the uncorrected α level (as indicated by asterisks in Figure 3). More information on centrality indices that were significantly different can be found in Table 4 (all, including nonsignificant, differences between centrality indices can be found in Supplement E of Multimedia Appendix 1).

Figure 3. The standardized centrality indices out-strength and in-strength among ecological momentary assessment and physical activity data within the networks of the no-lockdown and lockdown stages. The statistically significant indices (permutation tests using a two-sided P value at the uncorrected α level) are marked with asterisks.

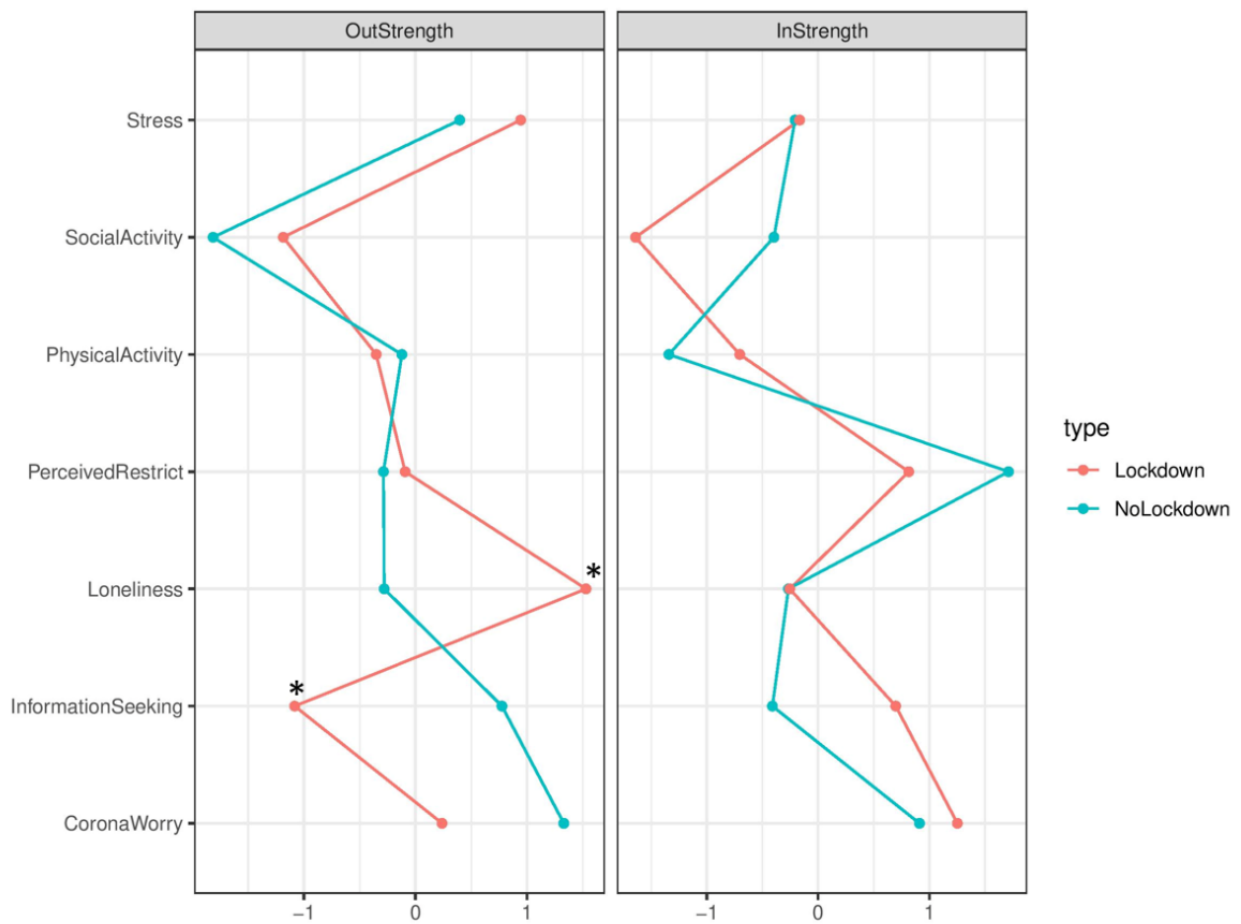


Table 4. Significant differences in variable out-strength between lockdown and no-lockdown stages.

Variable	Out-strength		Difference	P value
	No-lockdown	Lockdown		
Information-seeking	0.3129	0.1677	0.1452	.03
Loneliness	0.4109	0.6084	-0.1975	.04

Discussion

Principal Findings

The COVID-19 pandemic increased mental health problems worldwide [2,61]. Our study sheds light on the mechanisms

with which a lockdown affects mental health during the COVID-19 pandemic. Compared to no-lockdown, during lockdown, loneliness had a stronger impact on pandemic-related cognitions and behaviors such as perceived restrictions and worries about the pandemic. In turn, pandemic-related cognitions

and behaviors reinforced each other and increased stress across lockdown stages. Finally, we found engaging in daily physical activity to be an effective strategy against feelings of loneliness during lockdown. In sum, our results suggest that when strict lockdown measures are in place, loneliness is the central trigger of stress-related behaviors and cognitions. Thus, loneliness should be prioritized in mental health interventions in the context of pandemic-related psychological distress.

Loneliness is a distressing emotional state in which one experiences a discrepancy between the desired and perceived quantity and quality of social relations [62]. Previous studies showed that lonely individuals exhibit a negative information bias such as increased attention for social threatening stimuli, negative and hostile intent attributions, expectation of rejection, and rumination [63]. We found that during lockdown, feelings of loneliness had the highest out-strength, indicating that loneliness is the central trigger of stress-related behaviors and cognitions. Compared to a no-lockdown, a lockdown increased the out-strength of loneliness, which indicates that loneliness has a more central role in affecting stress-related cognitions and behaviors during lockdown. Moreover, during lockdown, the influence of loneliness on perceptions of restriction and COVID-19-related worry increased. Thus, a lockdown changes the way loneliness interacts with pandemic-related behaviors and cognitions.

COVID-19-related-worries, feelings of restriction, and information-seeking were mutually reinforcing over time in both the no-lockdown and lockdown stages, resulting in a vicious stress-inducing cycle from which it can be increasingly difficult to escape. Information-seeking had less out-strength during lockdown compared to the no-lockdown stage, which indicates that COVID-19-related information-seeking has a more central role during a no-lockdown period. During lockdown, information-seeking at one moment led to less information-seeking at the next moment (ie, weaker autocorrelation), and its influence on perceived restrictions and COVID-19-related worry decreased. These findings contrast earlier reports concluding a more significant influence of information-seeking during lockdown, based on findings of increased averaged information-seeking [19,21]. Moreover, during the no-lockdown stage, “feeling restricted” increased information-seeking, whereas during lockdown, “feeling restricted” decreased information-seeking. This suggests that during a no-lockdown stage, people are in a type of information-approach state, whereas during lockdown, people are more likely to be in an information-avoidance state. Therefore, the best moment to communicate COVID-19-relevant information such as safety behaviors might be an early pandemic stage when no lockdown measures are in place.

Physical activity increased social activity in both the no-lockdown and lockdown stages. This association might result from public health recommendations that suggest meeting people only outside enclosed spaces. During COVID-19, people might have combined physical and social activity (ie, they found a companion to go for a walk or hike outside). Physical activity can also help to form interpersonal relationships (eg, attending a virtual group fitness class). Moreover, physical activity

decreased feelings of loneliness during lockdown. A possible reason is that physical activity can mediate contextual influences on loneliness (eg, being in nature and physically active rather than sitting at home and leading a sedentary lifestyle) [64]. Meeting more people did not decrease feelings of loneliness in either of the lockdown stages. A potential explanation is that feelings of loneliness are not caused by the number of social contacts but rather the perception that current relationships do not match desired relationships (eg, the other person being attentive to one’s needs) [65]. Finally, physical activity and social activity were associated with decreased stress only during the lockdown stage, indicating that during lockdown, these stress-buffering behaviors become effective.

Perspectives on Mental Health Interventions

We found that loneliness has the highest temporal effect on all measured moment-to-moment pandemic-related cognition and behaviors during lockdown. This, in turn, suggests that loneliness can be a central trigger of stress-related behaviors and cognitions. Our study suggests that mental health interventions during the pandemic lockdown should prioritize the feeling of loneliness rather than pandemic-related rumination, feelings of restriction, or information-seeking. This could be achieved by a digital mental health approach (eg, online therapy or smartphone-based interventions) that fosters a sense of belonging and community [66-70]. To our knowledge, this is the first study to use a temporal network model comparison approach to identify and refine mental health intervention targets. This approach might be valuable to identify possible temporal causal trigger variables for negative cognitions and behaviors in other types of mental health interventions as well.

Limitations

This was a natural experiment with high ecological validity but low control for extraneous variables, including seasonal effects [71]. Moreover, we cannot exclude the possibility that the observed interactions are influenced by other unmeasured underlying factors [72]. In addition, we have independent samples for comparisons of lockdown and no-lockdown stages. Thus, we cannot exclude the possibility that differences in sample characteristics may have influenced the results. However, except for the loneliness score and gender distribution, the samples did not differ in any of the measured variables. We assume that the slightly higher loneliness measure (ULS-8) in the lockdown sample was due to the lockdown. However, it cannot be ruled out that we recruited participants who were generally lonelier in the lockdown sample by chance. Gender did not have an influence on any of the measured variables, except for time spent on social activities. Here, women reported higher values than men or diverse genders. Taken together, it is unlikely that there is a major bias in our central findings due to differences in sample characteristics.

Conclusion

To develop effective pandemic mental health interventions, it is crucial to understand the temporal dynamics of mental health factors during a COVID-19 lockdown. In comparison to a no-lockdown stage, a lockdown increased the central role of loneliness in triggering pandemic-related behaviors and

cognition. In turn, pandemic-related cognitions and behaviors such as perceived restrictions and worries about the pandemic reinforced each other and increased stress. In addition, we found that physical activity can be an effective buffer against stress

and loneliness during lockdown. Our results suggest that loneliness can be the central trigger for stress-related behaviors and cognitions during lockdown and therefore should be prioritized in mental health interventions.

Acknowledgments

This research was supported by the Berlin University Alliance (grant to SL and SH) and the Elsa-Neumann Stipendium (grant to MH).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplemental background, analysis methods, and data (Supplement A-G).

[DOCX File, 31 KB - [jmir_v24i3e32598_app1.docx](#)]

References

1. WHO coronavirus (COVID-19) dashboard. World Health Organization. URL: <https://covid19.who.int/> [accessed 2021-11-28]
2. 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](#)]
3. 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](#)]
4. O'Connor RC, Wetherall K, Cleare S, McClelland H, Melson AJ, Niedzwiedz CL, et al. *Br J Psychiatry* 2021 Jun;218(6):326-333 [FREE Full text] [doi: [10.1192/bjp.2020.212](#)] [Medline: [33081860](#)]
5. Ozamiz-Etxebarria N, Idoiaga Mondragon N, Dosil Santamaría M, Picaza Gorrotategi M. Psychological symptoms during the two stages of lockdown in response to the COVID-19 outbreak: an investigation in a sample of citizens in northern Spain. *Front Psychol* 2020 Jun 18;11:1491. [doi: [10.3389/fpsyg.2020.01491](#)]
6. Nelson BW, Pettitt A, Flannery JE, Allen NB. Rapid assessment of psychological and epidemiological correlates of COVID-19 concern, financial strain, and health-related behavior change in a large online sample. *PLoS One* 2020 Nov 11;15(11):e0241990 [FREE Full text] [doi: [10.1371/journal.pone.0241990](#)] [Medline: [33175882](#)]
7. Liu S, Heinzel S, Haucke MN, Heinz A. Increased psychological distress, loneliness, and unemployment in the spread of COVID-19 over 6 months in Germany. *Medicina* 2021 Jan 09;57(1):53 [FREE Full text] [doi: [10.3390/medicina57010053](#)] [Medline: [33435368](#)]
8. Wichers M, Wigman JTW, Myin-Germeys I. Micro-level affect dynamics in psychopathology viewed from complex dynamical system theory. *Emotion Review* 2015 Jul 09;7(4):362-367. [doi: [10.1177/1754073915590623](#)]
9. Bringmann LF, Elmer T, Epskamp S, Krause RW, Schoch D, Wichers M, et al. What do centrality measures measure in psychological networks? *J Abnorm Psychol* 2019 Nov;128(8):892-903. [doi: [10.1037/abn0000446](#)] [Medline: [31318245](#)]
10. van Gils A, Burton C, Bos EH, Janssens KA, Schoevers RA, Rosmalen JG. Individual variation in temporal relationships between stress and functional somatic symptoms. *J Psychosom Res* 2014 Jul;77(1):34-39. [doi: [10.1016/j.jpsychores.2014.04.006](#)] [Medline: [24913339](#)]
11. Deng W, Cheung S, Tsao S, Wang X, Tiwari A. Telomerase activity and its association with psychological stress, mental disorders, lifestyle factors and interventions: A systematic review. *Psychoneuroendocrinology* 2016 Feb;64:150-163. [doi: [10.1016/j.psyneuen.2015.11.017](#)] [Medline: [26677763](#)]
12. Heinz A, Deserno L, Reininghaus U. Urbanicity, social adversity and psychosis. *World Psychiatry* 2013 Oct 04;12(3):187-197. [doi: [10.1002/wps.20056](#)] [Medline: [24096775](#)]
13. Narita Z, Stickley A, DeVlyder J. Loneliness and psychotic experiences in a general population sample. *Schizophr Res* 2020 Apr;218:146-150. [doi: [10.1016/j.schres.2020.01.018](#)] [Medline: [32014362](#)]
14. Beutel ME, Klein EM, Brähler E, Reiner I, Jünger C, Michal M, et al. Loneliness in the general population: prevalence, determinants and relations to mental health. *BMC Psychiatry* 2017 Mar 20;17(1):97 [FREE Full text] [doi: [10.1186/s12888-017-1262-x](#)] [Medline: [28320380](#)]
15. Heinz AJ, Beck A, Meyer-Lindenberg A, Sterzer P, Heinz A. Cognitive and neurobiological mechanisms of alcohol-related aggression. *Nat Rev Neurosci* 2011 Jun 02;12(7):400-413. [doi: [10.1038/nrn3042](#)] [Medline: [21633380](#)]

16. Rauschenberg C, Schick A, Goetzl C, Roehr S, Riedel-Heller SG, Koppe G, et al. Social isolation, mental health, and use of digital interventions in youth during the COVID-19 pandemic: A nationally representative survey. *Eur Psychiatry* 2021 Mar 09;64(1):e20 [[FREE Full text](#)] [doi: [10.1192/j.eurpsy.2021.17](https://doi.org/10.1192/j.eurpsy.2021.17)] [Medline: [33686930](#)]
17. 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 May 06;23(5):e21864 [[FREE Full text](#)] [doi: [10.2196/21864](https://doi.org/10.2196/21864)] [Medline: [33891557](#)]
18. Zhang W, Yang X, Zhao J, Yang F, Jia Y, Cui C, et al. Depression and psychological-behavioral responses among the general public in China during the early stages of the COVID-19 pandemic: survey study. *J Med Internet Res* 2020 Sep 04;22(9):e22227 [[FREE Full text](#)] [doi: [10.2196/22227](https://doi.org/10.2196/22227)] [Medline: [32886066](#)]
19. Terry PC, Parsons-Smith RL, Terry VR. Mood responses associated with COVID-19 restrictions. *Front Psychol* 2020 Nov 4;11:589598. [doi: [10.3389/fpsyg.2020.589598](https://doi.org/10.3389/fpsyg.2020.589598)] [Medline: [33312153](#)]
20. Kleiman EM, Yeager AL, Grove JL, Kellerman JK, Kim JS. Real-time mental health impact of the COVID-19 pandemic on college students: ecological momentary assessment study. *JMIR Ment Health* 2020 Dec 15;7(12):e24815 [[FREE Full text](#)] [doi: [10.2196/24815](https://doi.org/10.2196/24815)] [Medline: [33207308](#)]
21. 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](https://doi.org/10.2196/20185)] [Medline: [32519963](#)]
22. Arend A, Blechert J, Pannicke B, Reichenberger J. Increased screen use on days with increased perceived COVID-19-related confinements—a day level ecological momentary assessment study. *Front Public Health* 2020 Feb 2;8:623205. [doi: [10.3389/fpubh.2020.623205](https://doi.org/10.3389/fpubh.2020.623205)] [Medline: [33634062](#)]
23. Jeste DV, Lee EE, Cacioppo S. Battling the modern behavioral epidemic of loneliness: suggestions for research and interventions. *JAMA Psychiatry* 2020 Jun 01;77(6):553-554 [[FREE Full text](#)] [doi: [10.1001/jamapsychiatry.2020.0027](https://doi.org/10.1001/jamapsychiatry.2020.0027)] [Medline: [32129811](#)]
24. Cacioppo J, Cacioppo S. Loneliness in the modern age: An evolutionary theory of loneliness (ETL). *Adv Exp Psychol* 2018;58:127-197. [doi: [10.1016/bs.aesp.2018.03.003](https://doi.org/10.1016/bs.aesp.2018.03.003)]
25. Haucke M, Liu S, Heinzel S. The persistence of the impact of COVID-19-related distress, mood inertia, and loneliness on mental health during a postlockdown period in Germany: an ecological momentary assessment study. *JMIR Ment Health* 2021 Aug 26;8(8):e29419 [[FREE Full text](#)] [doi: [10.2196/29419](https://doi.org/10.2196/29419)] [Medline: [34347622](#)]
26. Singer C. Health effects of social isolation and loneliness. *J Aging Life Care* 2018;28(1):4-8.
27. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. *Lancet* 2020 Mar 14;395(10227):912-920 [[FREE Full text](#)] [doi: [10.1016/S0140-6736\(20\)30460-8](https://doi.org/10.1016/S0140-6736(20)30460-8)] [Medline: [32112714](#)]
28. Heinz A, Zhao X, Liu S. Implications of the association of social exclusion with mental health. *JAMA Psychiatry* 2020 Feb 01;77(2):113-114. [doi: [10.1001/jamapsychiatry.2019.3009](https://doi.org/10.1001/jamapsychiatry.2019.3009)] [Medline: [31596429](#)]
29. Burg MM, Schwartz JE, Kronish IM, Diaz KM, Alcantara C, Duer-Hefele J, et al. Does stress result in you exercising less? Or does exercising result in you being less stressed? Or is it both? Testing the bi-directional stress-exercise association at the group and person (N of 1) level. *Ann Behav Med* 2017 Dec 13;51(6):799-809 [[FREE Full text](#)] [doi: [10.1007/s12160-017-9902-4](https://doi.org/10.1007/s12160-017-9902-4)] [Medline: [28290065](#)]
30. Hillman CH, Erickson KI, Kramer AF. Be smart, exercise your heart: exercise effects on brain and cognition. *Nat Rev Neurosci* 2008 Jan;9(1):58-65. [doi: [10.1038/nrn2298](https://doi.org/10.1038/nrn2298)] [Medline: [18094706](#)]
31. Hamilton M, Hamilton D, Zderic T. Sedentary behavior as a mediator of type 2 diabetes. *Med Sport Sci* 2014;60:11-26 [[FREE Full text](#)] [doi: [10.1159/000357332](https://doi.org/10.1159/000357332)] [Medline: [25226797](#)]
32. Myers J, McAuley P, Lavie CJ, Despres J, Arena R, Kokkinos P. Physical activity and cardiorespiratory fitness as major markers of cardiovascular risk: their independent and interwoven importance to health status. *Prog Cardiovasc Dis* 2015 Jan;57(4):306-314. [doi: [10.1016/j.pcad.2014.09.011](https://doi.org/10.1016/j.pcad.2014.09.011)] [Medline: [25269064](#)]
33. Moore SC, Lee I, Weiderpass E, Campbell PT, Sampson JN, Kitahara CM, et al. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. *JAMA Intern Med* 2016 Jun 01;176(6):816-825 [[FREE Full text](#)] [doi: [10.1001/jamainternmed.2016.1548](https://doi.org/10.1001/jamainternmed.2016.1548)] [Medline: [27183032](#)]
34. Vancampfort D, Firth J, Schuch FB, Rosenbaum S, Mugisha J, Hallgren M, et al. Sedentary behavior and physical activity levels in people with schizophrenia, bipolar disorder and major depressive disorder: a global systematic review and meta-analysis. *World Psychiatry* 2017 Oct 21;16(3):308-315. [doi: [10.1002/wps.20458](https://doi.org/10.1002/wps.20458)] [Medline: [28941119](#)]
35. Fox KR. The influence of physical activity on mental well-being. *Public Health Nutr* 2007 Jan 02;2(3a):411-418. [doi: [10.1017/S1368980099000567](https://doi.org/10.1017/S1368980099000567)]
36. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. *Ann Intern Med* 2015 Jan 20;162(2):123-132. [doi: [10.7326/M14-1651](https://doi.org/10.7326/M14-1651)] [Medline: [25599350](#)]
37. Chau JY, Grunseit AC, Chey T, Stamatakis E, Brown WJ, Matthews CE, et al. Daily sitting time and all-cause mortality: a meta-analysis. *PLoS One* 2013 Nov 13;8(11):e80000 [[FREE Full text](#)] [doi: [10.1371/journal.pone.0080000](https://doi.org/10.1371/journal.pone.0080000)] [Medline: [24236168](#)]

38. Wilmot EG, Edwardson CL, Achana FA, Davies MJ, Gorely T, Gray LJ, et al. Sedentary time in adults and the association with diabetes, cardiovascular disease and death: systematic review and meta-analysis. *Diabetologia* 2012 Nov;55(11):2895-2905. [doi: [10.1007/s00125-012-2677-z](https://doi.org/10.1007/s00125-012-2677-z)] [Medline: [22890825](https://pubmed.ncbi.nlm.nih.gov/22890825/)]
39. Kandola A, Ashdown-Franks G, Hendrikse J, Sabiston CM, Stubbs B. Physical activity and depression: towards understanding the antidepressant mechanisms of physical activity. *Neurosci Biobehav Rev* 2019 Dec;107:525-539. [doi: [10.1016/j.neubiorev.2019.09.040](https://doi.org/10.1016/j.neubiorev.2019.09.040)] [Medline: [31586447](https://pubmed.ncbi.nlm.nih.gov/31586447/)]
40. Sani SHZ, Fathirezaie Z, Brand S, Pühse U, Holsboer-Trachsler E, Gerber M, et al. Physical activity and self-esteem: testing direct and indirect relationships associated with psychological and physical mechanisms. *Neuropsychiatr Dis Treat* 2016 Oct;12:2617-2625. [doi: [10.2147/ndt.s116811](https://doi.org/10.2147/ndt.s116811)]
41. Rowlands A, Fraysse F, Catt M. Comparability of measured acceleration from accelerometry-based activity monitors. *Med Sci Sports Exerc* Jan 2015;47(1):201-210. [doi: [10.1249/mss.0000000000000394](https://doi.org/10.1249/mss.0000000000000394)]
42. van Hees VT, Renström F, Wright A, Gradmark A, Catt M, Chen KY, et al. Estimation of daily energy expenditure in pregnant and non-pregnant women using a wrist-worn tri-axial accelerometer. *PLoS One* 2011 Jul 29;6(7):e22922 [FREE Full text] [doi: [10.1371/journal.pone.0022922](https://doi.org/10.1371/journal.pone.0022922)] [Medline: [21829556](https://pubmed.ncbi.nlm.nih.gov/21829556/)]
43. Liu S, Heinz A. Cross-cultural validity of psychological distress measurement during the coronavirus pandemic. *Pharmacopsychiatry* 2020 Sep 24;53(5):237-238. [doi: [10.1055/a-1190-5029](https://doi.org/10.1055/a-1190-5029)] [Medline: [32583390](https://pubmed.ncbi.nlm.nih.gov/32583390/)]
44. Hays R, DiMatteo MR. A short-form measure of loneliness. *J Pers Assess* 1987 Mar 1;51(1):69-81. [doi: [10.1207/s15327752jpa5101_6](https://doi.org/10.1207/s15327752jpa5101_6)] [Medline: [3572711](https://pubmed.ncbi.nlm.nih.gov/3572711/)]
45. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. *Annu Rev Clin Psychol* 2008 Apr;4(1):1-32. [doi: [10.1146/annurev.clinpsy.3.022806.091415](https://doi.org/10.1146/annurev.clinpsy.3.022806.091415)] [Medline: [18509902](https://pubmed.ncbi.nlm.nih.gov/18509902/)]
46. Hamilton JD. Time series analysis. Princeton, NJ: Princeton University Press; 1994:306-308.
47. van Hees VT, Gorzelniak L, Dean León EC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS One* 2013 Apr 23;8(4):e61691 [FREE Full text] [doi: [10.1371/journal.pone.0061691](https://doi.org/10.1371/journal.pone.0061691)] [Medline: [23626718](https://pubmed.ncbi.nlm.nih.gov/23626718/)]
48. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva ICM, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J Appl Physiol* 2014 Oct 01;117(7):738-744 [FREE Full text] [doi: [10.1152/jappphysiol.00421.2014](https://doi.org/10.1152/jappphysiol.00421.2014)] [Medline: [25103964](https://pubmed.ncbi.nlm.nih.gov/25103964/)]
49. Liu H, Han F, Yuan M, Lafferty J, Wasserman L. High-dimensional semiparametric Gaussian copula graphical models. *Ann Statist* 2012 Aug 1;40(4):2293-2326. [doi: [10.1214/12-aos1037](https://doi.org/10.1214/12-aos1037)]
50. de Haan-Rietdijk S, Kuppens P, Hamaker EL. What's in a day? A guide to decomposing the variance in intensive longitudinal data. *Front Psychol* 2016 Jun 14;7:891. [doi: [10.3389/fpsyg.2016.00891](https://doi.org/10.3389/fpsyg.2016.00891)] [Medline: [27378986](https://pubmed.ncbi.nlm.nih.gov/27378986/)]
51. R code for the statistical analysis. OSF. URL: <https://osf.io/zskgm/> [accessed 2022-03-12]
52. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: network visualizations of relationships in psychometric data. *J Stat Soft* 2012;48(4):1-18. [doi: [10.18637/jss.v048.i04](https://doi.org/10.18637/jss.v048.i04)]
53. Snippe E, Viechtbauer W, Geschwind N, Klippel A, de Jonge P, Wichers M. The impact of treatments for depression on the dynamic network structure of mental states: two randomized controlled trials. *Sci Rep* 2017 Apr 20;7(1):46523. [doi: [10.1038/srep46523](https://doi.org/10.1038/srep46523)] [Medline: [28425449](https://pubmed.ncbi.nlm.nih.gov/28425449/)]
54. Klippel A, Viechtbauer W, Reininghaus U, Wigman J, van Borkulo C, MERGE, et al. The cascade of stress: a network approach to explore differential dynamics in populations varying in risk for psychosis. *Schizophr Bull* 2018 Feb 15;44(2):328-337 [FREE Full text] [doi: [10.1093/schbul/sbx037](https://doi.org/10.1093/schbul/sbx037)] [Medline: [28338969](https://pubmed.ncbi.nlm.nih.gov/28338969/)]
55. Groen RN, Snippe E, Bringmann LF, Simons CJ, Hartmann JA, Bos EH, et al. Capturing the risk of persisting depressive symptoms: a dynamic network investigation of patients' daily symptom experiences. *Psychiatry Res* 2019 Jan;271:640-648. [doi: [10.1016/j.psychres.2018.12.054](https://doi.org/10.1016/j.psychres.2018.12.054)] [Medline: [30791337](https://pubmed.ncbi.nlm.nih.gov/30791337/)]
56. Bringmann LF, Elmer T, Epskamp S, Krause RW, Schoch D, Wichers M, et al. What do centrality measures measure in psychological networks? *J Abnorm Psychol* 2019 Nov;128(8):892-903. [doi: [10.1037/abn0000446](https://doi.org/10.1037/abn0000446)] [Medline: [31318245](https://pubmed.ncbi.nlm.nih.gov/31318245/)]
57. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: generalizing degree and shortest paths. *Social Networks* 2010 Jul;32(3):245-251. [doi: [10.1016/j.socnet.2010.03.006](https://doi.org/10.1016/j.socnet.2010.03.006)]
58. Bringmann LF, Vissers N, Wichers M, Geschwind N, Kuppens P, Peeters F, et al. A network approach to psychopathology: new insights into clinical longitudinal data. *PLoS One* 2013 Apr 4;8(4):e60188 [FREE Full text] [doi: [10.1371/journal.pone.0060188](https://doi.org/10.1371/journal.pone.0060188)] [Medline: [23593171](https://pubmed.ncbi.nlm.nih.gov/23593171/)]
59. Epskamp S, Waldorp LJ, Möttus R, Borsboom D. The Gaussian graphical model in cross-sectional and time-series data. *Multivariate Behav Res* 2018 Apr 16;53(4):453-480. [doi: [10.1080/00273171.2018.1454823](https://doi.org/10.1080/00273171.2018.1454823)] [Medline: [29658809](https://pubmed.ncbi.nlm.nih.gov/29658809/)]
60. Fried E, Papanikolaou F, Epskamp S. Mental health and social contact during the COVID-19 pandemic: an ecological momentary assessment study. *Clin Psychol Sci* 2021 Jun 07;10(2):340-354. [doi: [10.1177/21677026211017839](https://doi.org/10.1177/21677026211017839)]
61. Reger MA, Stanley IH, Joiner TE. Suicide mortality and coronavirus disease 2019—a perfect storm? *JAMA Psychiatry* 2020 Nov 01;77(11):1093-1094. [doi: [10.1001/jamapsychiatry.2020.1060](https://doi.org/10.1001/jamapsychiatry.2020.1060)] [Medline: [32275300](https://pubmed.ncbi.nlm.nih.gov/32275300/)]
62. Cacioppo JT, Hawkley LC, Ernst JM, Burleson M, Bertson GG, Nouriani B, et al. Loneliness within a nomological net: an evolutionary perspective. *J Res Person* 2006 Dec;40(6):1054-1085. [doi: [10.1016/j.jrp.2005.11.007](https://doi.org/10.1016/j.jrp.2005.11.007)]

63. Borawski D. Authenticity and rumination mediate the relationship between loneliness and well-being. *Curr Psychol* 2019 Aug 13;40(9):4663-4672. [doi: [10.1007/s12144-019-00412-9](https://doi.org/10.1007/s12144-019-00412-9)]
64. Pels F, Kleinert J. Loneliness and physical activity: a systematic review. *Int Rev Sport Exerc Psychol* 2016 Jun 07;9(1):231-260. [doi: [10.1080/1750984x.2016.1177849](https://doi.org/10.1080/1750984x.2016.1177849)]
65. Russell DW, Cutrona CE, McRae C, Gomez M. Is loneliness the same as being alone? *J Psychol* 2012 Jan;146(1-2):7-22. [doi: [10.1080/00223980.2011.589414](https://doi.org/10.1080/00223980.2011.589414)] [Medline: [22303609](https://pubmed.ncbi.nlm.nih.gov/22303609/)]
66. Torous J, Wykes T. Opportunities From the coronavirus disease 2019 pandemic for transforming psychiatric care with telehealth. *JAMA Psychiatry* 2020 Dec 01;77(12):1205-1206. [doi: [10.1001/jamapsychiatry.2020.1640](https://doi.org/10.1001/jamapsychiatry.2020.1640)] [Medline: [32391857](https://pubmed.ncbi.nlm.nih.gov/32391857/)]
67. Reinhardt I, Gouzoulis-Mayfrank E, Zielasek J. Use of telepsychiatry in emergency and crisis intervention: current evidence. *Curr Psychiatry Rep* 2019 Jul 01;21(8):63. [doi: [10.1007/s11920-019-1054-8](https://doi.org/10.1007/s11920-019-1054-8)] [Medline: [31263972](https://pubmed.ncbi.nlm.nih.gov/31263972/)]
68. 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](https://doi.org/10.2196/18848)] [Medline: [32213476](https://pubmed.ncbi.nlm.nih.gov/32213476/)]
69. Cohen J, Torous J. The potential of object-relations theory for improving engagement with health apps. *JAMA* 2019 Dec 10;322(22):2169-2170. [doi: [10.1001/jama.2019.17141](https://doi.org/10.1001/jama.2019.17141)] [Medline: [31725854](https://pubmed.ncbi.nlm.nih.gov/31725854/)]
70. 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](https://doi.org/10.2196/24268)] [Medline: [33264099](https://pubmed.ncbi.nlm.nih.gov/33264099/)]
71. Shaughnessy J, Zechmeister E, Zechmeister J. *Research methods in psychology*. New York, NY: McGraw-Hill; 2000.
72. Wichers M, Wigman JTW, Bringmann LF, de Jonge P. Mental disorders as networks: some cautionary reflections on a promising approach. *Soc Psychiatry Psychiatr Epidemiol* 2017 Feb 16;52(2):143-145. [doi: [10.1007/s00127-016-1335-z](https://doi.org/10.1007/s00127-016-1335-z)] [Medline: [28210750](https://pubmed.ncbi.nlm.nih.gov/28210750/)]

Abbreviations

- CPDI:** COVID-19 Peritraumatic Distress Index
EMA: ecological momentary assessment
ENMO: Euclidean norm minus one
KPSS: Kwiatkowski-Phillips-Schmidt-Shin
ULS-8: University of California Los Angeles Loneliness Scale
VAR: vector autoregressive model

Edited by C Basch; submitted 03.08.21; peer-reviewed by M Auriacombe, R Menhas; comments to author 07.11.21; revised version received 29.11.21; accepted 12.12.21; published 17.03.22.

Please cite as:

Haucke M, Heinz A, Liu S, Heinzl S

The Impact of COVID-19 Lockdown on Daily Activities, Cognitions, and Stress in a Lonely and Distressed Population: Temporal Dynamic Network Analysis

J Med Internet Res 2022;24(3):e32598

URL: <https://www.jmir.org/2022/3/e32598>

doi: [10.2196/32598](https://doi.org/10.2196/32598)

PMID: [35191843](https://pubmed.ncbi.nlm.nih.gov/35191843/)

©Matthias Haucke, Andreas Heinz, Shuyan Liu, Stephan Heinzl. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 17.03.2022. 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.

Original Paper

Stratified Impacts of the Infodemic During the COVID-19 Pandemic: Cross-sectional Survey in 6 Asian Jurisdictions

Xi Chen¹, PhD; Fen Lin², PhD; Edmund W Cheng³, PhD

¹The Jockey Club School of Public Health and Primary Care, The Chinese University of Hong Kong, Shatin, Hong Kong

²Department of Media and Communication, City University of Hong Kong, Kowloon Tong, Hong Kong

³Department of Public Policy, City University of Hong Kong, Kowloon Tong, Hong Kong

Corresponding Author:

Fen Lin, PhD

Department of Media and Communication

City University of Hong Kong

M5086, 5/F, Run Run Shaw Creative Media Centre

18 Tat Hong Avenue

Kowloon Tong

Hong Kong

Phone: 852 34428691

Email: fenlin@cityu.edu.hk

Abstract

Background: Although timely and accurate information during the COVID-19 pandemic is essential for containing the disease and reducing mental distress, an infodemic, which refers to an overabundance of information, may trigger unpleasant emotions and reduce compliance. Prior research has shown the negative consequences of an infodemic during the pandemic; however, we know less about which subpopulations are more exposed to the infodemic and are more vulnerable to the adverse psychological and behavioral effects.

Objective: This study aimed to examine how sociodemographic factors and information-seeking behaviors affect the perceived information overload during the COVID-19 pandemic. We also investigated the effect of perceived information overload on psychological distress and protective behavior and analyzed the socioeconomic differences in the effects.

Methods: The data for this study were obtained from a cross-national survey of residents in 6 jurisdictions in Asia in May 2020. The survey targeted residents aged 18 years or older. A probability-based quota sampling strategy was adopted to ensure that the selected samples matched the population's geographical and demographic characteristics released by the latest available census in each jurisdiction. The final sample included 10,063 respondents. Information overload about COVID-19 was measured by asking the respondents to what extent they feel overwhelmed by news related to COVID-19. The measure of psychological distress was adapted from the Posttraumatic Stress Disorder Checklist for Diagnostic and Statistical Manual of Mental Disorders 5 (DSM-5). Protective behaviors included personal hygienic behavior and compliance with social distancing measures.

Results: Younger respondents and women ($b=0.20$, 95% CI 0.14 to 0.26) were more likely to perceive information overload. Participants self-perceived as upper or upper-middle class ($b=0.19$, 95% CI 0.09 to 0.30) and those with full-time jobs ($b=0.11$, 95% CI 0.04 to 0.17) tended to perceive higher information overload. Respondents who more frequently sought COVID-19 information from newspapers ($b=0.12$, 95% CI 0.11 to 0.14), television ($b=0.07$, 95% CI 0.05 to 0.09), and family and friends ($b=0.11$, 95% CI 0.09 to 0.14) were more likely to feel overwhelmed. In contrast, obtaining COVID-19 information from online news outlets and social media was not associated with perceived information overload. There was a positive relationship between perceived information overload and psychological distress ($b=2.18$, 95% CI 2.09 to 2.26). Such an association was stronger among urban residents, full-time employees, and those living in privately owned housing. The effect of perceived information overload on protective behavior was not significant.

Conclusions: Our findings revealed that respondents who were younger, were female, had a higher socioeconomic status (SES), and had vulnerable populations in the household were more likely to feel overwhelmed by COVID-19 information. Perceived information overload tended to increase psychological distress, and people with higher SES were more vulnerable to this adverse psychological consequence. Effective policies and interventions should be promoted to target vulnerable populations who are more susceptible to the occurrence and negative psychological influence of perceived information overload.

KEYWORDS

infodemic; information overload; psychological distress; protective behavior; cross-national survey; Asia; COVID-19

Introduction

Background

The COVID-19 pandemic has posed unprecedented challenges to public health and daily life worldwide. A cluster of COVID-19 cases was first reported in December 2019 in Wuhan of Hubei Province in China. Due to the proximity and various links to China, COVID-19 badly hit Asia early on. By imposing strict public health measures, some countries in Asia and the Pacific had better performance in containing the spread of COVID-19 compared with the rest of the world. However, the outbreaks of the Delta variant in several Asian countries and regions, including India, Singapore, Taiwan, and Thailand [1], show the difficulty in suppressing COVID-19 even with the implementation of social distancing and the availability of vaccines. As of May 31, 2021, cumulative cases in Asia had reached over 51.07 million—nearly one-third of the total cases reported globally—and the number deaths had reached over 682,000 [2].

The unpredictable course of the pandemic along with prolonged social distancing can negatively affect people's mental health, regardless of exposure to the disease itself [3]. A study of the general population in Hong Kong between March 2020 and May 2020 revealed that almost two-thirds of the respondents reported depression or anxiety disorders and about one-quarter met the criteria for risk of psychosis [4]. Also, about 45% of South Korean residents experienced moderate or higher symptoms of depression, anxiety, or stress, as shown by a survey from March 2020 to June 2020 [5]. Similarly, a study of the general population in Taiwan in April 2020 found that 55.8% of the participants reported sleep disturbance and 10.8% reported having suicidal thoughts in the previous week [6].

In addition, the psychological impact of COVID-19 has been fueled by an “infodemic,” which refers to “an overabundance of information—some accurate and some not—that makes it hard for people to find trustworthy sources and reliable guidance when they need it” [7]. An infodemic may cause a feeling of information overload when the amount of information to which people are exposed exceeds the optimal level that they can process and understand effectively [8]. Such a feeling of information overload may worsen when the information contains contradictory or uncertain contents [9,10]. As COVID-19 is a sudden disease outbreak with unknown causes and unpredictable course, its related information involves a high level of uncertainty. For instance, the prevention guidance for COVID-19 is ambiguous and changing as experts and authorities present different perspectives, particularly in the early stage of the outbreak [11]. The complexity of COVID-19 information (eg, contains too much scientific jargon) also contributes to a perception of information overload, as processing COVID-19 information requires great cognitive resources.

Although timely and accurate information during the pandemic helps individuals develop adequate risk perceptions, take preventive measures, and reduce mental distress, an infodemic and information overload can trigger unpleasant emotions, cause confusion and distrust among people, and impede effective public health responses [12-14]. Prior research has revealed that health information overload was negatively associated with one's attitudes and willingness to use medical services. Health information overload may also hinder the dissemination of health knowledge and affect public health decision-making. For example, cancer information overload may increase individuals' susceptibility to cancer fatalism [14], reduce the willingness to undergo cancer screening [15], and decrease self-management of chronic diseases [16]. Information overload and its related psychological discomfort can lead to information avoidance [17], reduce motivation for information sharing [18], and reduce health behaviors [19]. During the COVID-19 pandemic, people have been exposed to a great deal of information, which not only is based on scientific evidence but also contains misinformation and rumors from unreliable sources. Exposure to too much ambiguous information may pose an obstacle to appropriate responses and undermine trust in health institutions and programs [20]. As such, researchers and public health authorities should be mindful of the causes and consequences of information overload on emergency responses.

To date, only a few studies have investigated the prevalence and consequences of perceived information overload during the COVID-19 pandemic, most of which discussed the effect of information overload on compliance with protective behaviors [21-25]. In light of the psychological and emotional impacts of COVID-19 information (eg, fatigue and anxiety) [26,27], scholars have suggested that “future studies would add more valuable insights if they aim to investigate psychological and emotional responses of information overload” [25]. Although information overload is a stress indicator, previous findings indicated that information overload measured different concepts of perceived stress and the scales for measuring perceived information overload and perceived stress do not overlap [28].

Besides a lack of research on the psychological consequences of information overload about COVID-19, most existing studies only examined the impact of information overload in the general population while ignoring the potentially stratified impact of information overload in different subpopulations. Moreover, the scope of previous studies on information overload is limited. A recent systematic review of health information overload pointed out that most research has been conducted in the United States, focusing on cancer information overload perceived by cancer patients and thus called for extending the scope to other health issues in different contexts [29].

To fill the gaps, this paper aimed to systematically examine the level, associated factors, and psychological and behavioral consequences of perceived information overload about COVID-19 and explore the sociodemographic variances in the

susceptibility to and impact of information overload. The data were collected from a large-scale, cross-sectional survey of 10,063 residents in 6 jurisdictions in Asia, including Hong Kong, Taiwan, Japan, South Korea, Singapore, and Thailand, in May 2020. We focused on 4 research questions: (1) Which segments of the population perceived higher levels of information overload during the COVID-19 pandemic? (2) How has information-seeking behavior (ie, the frequency of accessing and perceived trustworthiness of COVID-19 information from different sources) affected the perception of information overload? (3) Would perceived information overload negatively affect psychological well-being and preventative behavior during the pandemic? (4) If so, which subpopulations are more vulnerable to the psychological and behavioral consequences of information overload? The findings would help identify vulnerable groups who are more susceptible to information overload about COVID-19 and its psychological and behavioral consequences and thus contribute to a nuanced understanding of the correlates and consequences of an infodemic and information overload during the COVID-19 pandemic.

Prior Work

As information overload arises when there is much more information available than an individual's information processing capacity [30], people who are more attentive to information or have a lower level of cognitive capacity for dealing with relevant information tend to perceive higher levels of information overload. Previous studies have shown that older adults [31] and women [32-35] experienced greater information overload. Also, those who are less educated [31,36,37] and of lower socioeconomic status (SES) [38,39] were more likely to feel overwhelmed by information, which may be due to their lower capacity for processing information.

The empirical results of the relationship between media exposure and information overload were mixed [38,40,41]. Moreover, the effect of various media types on perceived information overload may be different [36]. During the COVID-19 pandemic, information overload was higher among individuals who used social media as a source for COVID-19 information [42,43], perhaps because misinformation and rumors tend to spread more rapidly on social media platforms [44,45]. Thus, we proposed the following hypotheses: Hypothesis 1 (H1) is that the frequency of obtaining COVID-19 information from traditional sources (eg, newspaper, television, and family and friends) is negatively associated with perceived information overload. Hypothesis 2 (H2) is that the frequency of obtaining COVID-19 information from social media is positively associated with perceived information overload.

Although access to timely and quality health information during outbreaks of infectious diseases can effectively contain the spread of diseases and reduce depressive and anxious feelings [46,47], an overabundance of information can make people feel powerless and anxious, experience information fatigue, and use heuristic rather than systematic information processing [17,48]. Prior research has documented the adverse impact of excessive media consumption on psychological well-being during infectious health outbreaks such as the 2014 Ebola outbreak and the swine flu pandemic [49,50]. Likewise, during the current

COVID-19 pandemic, people are likely to develop negative emotions when experiencing information overload [51]. Based on the broadly consistent findings of the negative effect of information overload on psychological well-being, we posited the following hypothesis: Hypothesis 3 (H3) is that perceived information overload is positively associated with psychological distress during the COVID-19 pandemic.

Researchers have also investigated the role of various information sources in psychological well-being and coping behavior during the COVID-19 pandemic. Previous studies in mainland China [52] and Taiwan [53] have revealed that exposure to social media was positively associated with mental distress during the COVID-19 pandemic. In contrast, no such relation was found with traditional media (mass and print media) [42]. We thus proposed the following hypotheses about the effect of information sources and psychological well-being during the COVID-19 pandemic: Hypothesis 4 (H4) is that the frequency of obtaining COVID-19 information from traditional sources (eg, newspaper, TV, family, and friends) has a limited effect on psychological well-being during the COVID-19 pandemic. Hypothesis 5 (H5) is that the frequency of obtaining COVID-19 information from social media is positively associated with psychological distress.

Despite the well-documented relationship between perceived information overload and psychological distress, limited studies have investigated the socioeconomic differences in the relationship between perceived information overload and psychological well-being. Prior research suggested that confusing and ambiguous information is especially problematic for those experiencing communication inequality, such as the lack of access to relevant health information or the ability to make sense of information [54]. Populations at risk of communication inequality disproportionately consist of older adults, people with low educational levels and incomes, ethnic minorities, and residents of rural areas [54,55]. Thus, structurally vulnerable populations with communication disadvantages may experience higher levels of psychological distress because of perceived information overload. Therefore, we hypothesized the following: Hypothesis 6 (H6) is that the effects of perceived information overload on psychological distress are stronger among older people. Hypothesis 7 (H7) is that the effects of perceived information overload on psychological distress are stronger among respondents with lower SES.

In addition to psychological responses to information overload, people's health behavior is also likely to be influenced by information overload. Previous studies on health information overload have consistently shown that those who perceive higher information overload are less likely to perform health behaviors [19,56-58]. However, there were mixed findings on behavioral consequences of information overload during the COVID-19 pandemic. A study of 225 participants in Finland showed that perceived information overload had a negative effect on protective health behavior (eg, self-isolation) during the COVID-19 pandemic [59]. Nevertheless, other studies found no direct impact of information overload about COVID-19 and intention to adopt protective behavior [25]. Some studies even found a positive relationship between information overload and the adoption of protective behavior [23,60]. Given the ambiguity

in the empirical evidence, we tentatively hypothesized a negative relationship between perceived information overload and compliance behaviors during the COVID-19 pandemic: Hypothesis 8 (H8) is that participants who perceived higher levels of information overload about COVID-19 were less likely to adopt protective behaviors.

Methods

Study Design and Data Collection

The data for this study were obtained from a cross-national survey of public attitudes and responses toward COVID-19 in 6 jurisdictions in East and Southeast Asia, including Hong Kong, Taiwan, Japan, South Korea, Singapore, and Thailand. The survey was conducted by a group of scholars at the City University of Hong Kong between May 11, 2020 and May 26, 2020. The 6 regions were selected due to their geographical proximity to mainland China, the original epicenter of the COVID-19 pandemic, and were hardly hit by the pandemic since late January 2020. In late March 2020, they all entered a second wave of the pandemic as more imported cases from Europe and the United States were detected. They are the Tiger economies characterized by relatively high economic development and governing capacity in Asia. Yet, they also vary in regime types, media development, the stringency of public health measures, and effectiveness in containing the pandemic. Thus, these cases provide an ideal mixture of similarities and differences to examine the impacts of information-seeking behavior, perceived information overload, and psychological well-being during the COVID-19 pandemic.

The surveys were completed using online panels provided by a globally recognized professional survey company. The company's online panels consist of an opt-in list of 56,000 to

1,440,000 individuals relative to the population size in the 6 jurisdictions surveyed in this study. Online panels have been used increasingly among psychological, social, and medical research [61-64]. Using a panel provider for online research can help obtain a representative sample of the required size and facilitate quick completions for time-sensitive projects [65], especially during the COVID-19 pandemic. We reported our methods in line with the Checklist for Reporting Results of Internet E-Surveys (CHERRIES) [66].

For this study, we requested nationally representative samples of around 2000 adults aged 18 years or older in each of the 6 jurisdictions. Age and gender sampling quotas were set to match the latest available census estimates for age and gender in each jurisdiction. Participants were invited through email messages with an embedded link. The panel provider continuously invited participants until the predetermined quota was met. To increase the response rate, participants would get modest monetary rewards upon completion of the survey. Participation was voluntary, and all responses were anonymous. Details of the survey method of this project can be found elsewhere [67].

We developed a questionnaire that includes questions on perceived information overload about COVID-19, information-seeking behavior, psychological well-being, and protective behavior during the pandemic. We conducted a pre-test of the survey questions and modified wordings based on the feedback from the pre-testers. The questionnaire was available in English, Chinese, Korean, Japanese, and Thai for participants from different jurisdictions. A total of 12,062 representative respondents was collected, with approximately 2000 individuals in each jurisdiction. Cases with incomplete information were excluded from the analysis. The final sample size was 10,063 (see [Table 1](#) for more details).

Table 1. Sample characteristics.

Variable	Full sample (N=10,063), n (%)	Hong Kong (n=1813), n (%)	Japan (n=1372), n (%)	Singapore (n=1681), n (%)	South Korea (n=1749), n (%)	Taiwan (n=1695), n (%)	Thailand (n=1753), n (%)
Age (years)							
18-29	2444 (24.29)	413 (22.78)	254 (18.51)	400 (23.80)	431 (24.64)	411 (24.25)	535 (30.52)
30-39	2441 (24.26)	425 (23.44)	304 (22.16)	389 (23.14)	441 (25.21)	408 (24.07)	474 (27.04)
40-49	2379 (23.64)	470 (25.92)	263 (19.17)	385 (22.90)	451 (25.79)	394 (23.24)	416 (23.73)
50-59	1970 (19.58)	392 (21.62)	348 (25.36)	346 (20.58)	273 (15.61)	357 (21.06)	254 (14.49)
≥60	829 (8.24)	113 (6.23)	203 (14.80)	161 (9.58)	153 (8.75)	125 (7.37)	74 (4.22)
Sex							
Male	5258 (52.25)	865 (47.71)	761 (55.47)	867 (51.58)	927 (53.00)	955 (56.34)	883 (50.37)
Female	4805 (47.75)	948 (52.29)	611 (44.53)	814 (48.42)	822 (47.00)	740 (43.66)	870 (49.63)
Education							
Secondary school or below	1690 (16.79)	518 (28.57)	316 (23.03)	232 (13.80)	169 (9.66)	194 (11.45)	389 (22.19)
College or above	8373 (83.21)	1295 (71.43)	1056 (76.97)	1449 (86.20)	1580 (90.34)	1501 (88.55)	1364 (77.81)
Area							
Urban	8147 (80.96)	1673 (92.28)	802 (58.45)	1648 (87.33)	1580 (90.34)	1475 (87.02)	1149 (65.54)
Rural	1916 (19.04)	140 (7.72)	570 (41.55)	213 (12.67)	169 (9.66)	220 (12.98)	604 (34.46)
Perceived social status							
Lower or lower-middle class	4034 (40.09)	947 (52.23)	658 (47.96)	523 (31.11)	867 (49.57)	632 (37.29)	407 (23.22)
Middle class	4915 (48.84)	628 (34.64)	523 (38.12)	958 (56.99)	719 (41.11)	827 (48.79)	1260 (71.88)
Upper or upper-middle class	1114 (11.07)	238 (13.13)	191 (13.92)	200 (11.9)	163 (9.32)	236 (13.92)	86 (4.91)
Employment status							
Working full time	6278 (62.39)	1345 (74.19)	725 (52.84)	1155 (68.71)	886 (50.66)	1253 (73.92)	914 (52.14)
Other	3785 (37.61)	468 (25.81)	647 (47.16)	526 (31.29)	863 (49.34)	442 (26.08)	839 (47.86)
Housing type							
Privately owned housing	6043 (60.05)	1017 (56.09)	893 (65.09)	1210 (71.98)	1152 (65.87)	1200 (70.80)	1310 (74.73)
Other	4020 (39.95)	796 (43.91)	479 (34.91)	471 (28.02)	597 (34.13)	495 (29.20)	443 (25.27)
Chronic illness							
Yes	1378 (13.69)	198 (10.92)	244 (17.78)	195 (11.60)	343 (19.61)	161 (9.50)	237 (13.52)
No	8685 (86.31)	1615 (89.08)	1128 (82.22)	1486 (88.40)	1406 (80.39)	1534 (90.50)	1516 (86.48)
Having pregnant women or older adults (>65 years old) in the household							
Yes	2958 (29.39)	535 (29.51)	383 (27.92)	378 (22.49)	377 (21.56)	620 (36.58)	665 (37.93)
No	7105 (70.61)	1278 (70.49)	989 (72.08)	1303 (77.51)	1372 (78.44)	1075 (63.42)	1088 (62.07)
Having children aged under 12 years in the household							
Yes	2975 (29.56)	533 (29.4)	254 (18.51)	489 (29.09)	376 (21.50)	490 (28.91)	833 (47.52)
No	7088 (70.44)	1280 (70.6)	1118 (81.49)	1192 (70.91)	1373 (78.50)	1205 (71.09)	920 (52.48)

Ethical Review

The study was approved by the Human Subject Ethics Committee of the City University of Hong Kong (Ref No: 8-2020-04-E295-18). All necessary participant consent was obtained.

Measures

Psychological distress was gauged by the Posttraumatic Stress Disorder Checklist for Diagnostic and Statistical Manual of Mental Disorders (DSM-5; PCL-5) [68]. Five items were selected according to their relevance to COVID-19, such as

“Having trouble falling or staying asleep” and “Having trouble concentrating.” The responses ranged from 0 = “Not at all” to 6 = “Very much.” The scores of the 5 items were summed to represent psychological distress, with higher values indicating higher levels of psychological distress. The theoretical score range of the variable was 0-30. The internal reliability of the scale was satisfactory (Cronbach $\alpha=0.80$).

Protective behavior was assessed by asking the respondents to rate on a 7-point Likert scale the level of compliance with 6 protective behaviors suggested by the government (eg, “kept a distance of at least 2 meters to other people” and “wore a mask in public space”) in the past week. Higher scores indicated higher levels of compliance.

Perceived information overload about COVID-19 was measured by the question, “To what extent do you feel overwhelmed by news related to the COVID-19 pandemic?” (0 = “Not at all” to 6 = “Very much”). This item was selected and adapted from the Perceived Information Overload scale [28] and the Cancer Information Overload scale [14,69].

Information-seeking behavior was assessed by the frequency of accessing COVID-19 information from the following platforms: newspapers, television, online news outlets, and social media (jurisdiction-specific examples were included in blanket statements based on the most popular social media tools in each jurisdiction; 0 = “Never” to 6 = “Very frequently”). We also controlled for perceived source credibility and time spent searching for COVID-19 information. Perceived source credibility was gauged by asking the respondents to rate how trustworthy they considered news about COVID-19 from the sources mentioned above on a 7-point Likert scale (1 = “Not at all trustworthy” to 7 = “Very trustworthy”). We also controlled for the average time (in hours) the respondents spent viewing information about COVID-19 each day.

We adjusted for demographic variables, such as age (18-29, 30-39, 40-49, 50-59, ≥ 60 years) and sex (male vs female). We also included a series of socioeconomic factors, including education (secondary school or below vs college or above), rural or urban residence, employment status (working full-time vs other), and housing type (privately owned housing vs other). We also assessed perceived social status by asking the respondents to declare their perceived social position (lower or lower-middle class, middle class, upper or upper-middle class). In addition, we controlled for the perceived threat of COVID-19 and level of worry about contracting COVID-19. To assess the impact of having a vulnerable household member, we asked whether the respondent lived with a pregnant woman or an adult older than 65 years and whether they had a child aged 12 years or below.

Statistical Analysis

Descriptive statistics are reported as mean and SD for normally distributed continuous variables or median and IQR in the case of skewed distributions (Table 1). Normality of the distribution of statistics was tested using skewness or kurtosis tests for normality. For the regression analysis, we first used ordinary least squares (OLS) regression to examine the effects of sociodemographic variables and information-seeking behavior

on perceived information overload (outcome variable). Next, OLS regressions with robust standard errors were carried out to examine the effects of perceived information overload and information-seeking behavior on 2 outcome variables (psychological distress and protective behavior) after adjusting for various sociodemographic variables. Lastly, 2-way interaction terms were computed between perceived information overload and each sociodemographic factor (ie, age, sex, education, employment, urban or rural residence, perceived social status, and housing types) to examine whether the effects of information overload on psychological distress (outcome variable) varied across different sociodemographic groups. As the main effect of perceived information overload on protective behavior was not significant, we did not further examine the sociodemographic differences in such an effect. We used graphical and numerical tests to verify OLS assumptions, and the results showed no evident violation of OLS assumptions (detailed information is included in Multimedia Appendix 1). To account for potential heteroscedasticity, we obtained robust standard errors in OLS models [70]. All the analyses were performed using Stata 15.0. Unstandardized coefficients with 95% CIs are reported. A *P* value of .05 was set as the level of statistical significance.

Results

Sample Characteristics

Table 1 displays the background characteristics of the respondents. In the full sample, about one-half (4885/10,063, 48.54%) of the respondents were under 40 years old. There were slightly more men (5258/10,063, 52.25%) than women. Most respondents had a college education or above (8373/10,063, 83.21%), lived in urban areas (8147/10,063, 80.96%), worked full-time (6278/10,063, 62.39%), and lived in privately owned housing (6043/10,063, 60.05%). About 40% (4034/10,063, 40.09%) of respondents perceived themselves as lower or lower-middle class, while about 49% (4915/10,063, 48.84%) and 11% (1114/10,063, 11.07%) of respondents perceived their status as middle class or upper or upper-middle class, respectively. About 14% (1378/10,063, 13.69%) of respondents had a chronic illness. About one-third (2958/10,063, 29.39%) of respondents had family members vulnerable to COVID-19 (eg, pregnant women, older adults aged over 65 years, children under 12 years old at home).

Descriptive Statistics

Table 2 presents the description of key variables. Of a theoretical score range of 0 to 35, the mean psychological distress score among the full sample was 14.27 (SD 8.00). Respondents in South Korea reported the highest level of distress (mean 16.21), followed by residents in Hong Kong (mean 14.90), Thailand (mean 14.48), Singapore (mean 13.74), and Japan (mean 13.14). Taiwan residents reported the lowest level of psychological distress (mean 12.83). As for involvement in protective behaviors, the median for the full sample was 6 (out of a range of 1 to 7), with respondents in Taiwan reporting the lowest level of compliance (median 5.55) and respondents in Singapore reporting the highest compliance (median 6.50). The median level of perceived information overload was 3 (0-6; mean 3.35,

SD 1.64). Respondents in South Korea (mean 3.96) and Taiwan (mean 2.65) experienced the highest and lowest levels of information overload, respectively. Residents in Japan, Hong Kong, Singapore, and Thailand reported levels of information overload between 3.02 and 3.76. As for media use, respondents were less likely to use newspapers for COVID-19 information (median 3) and more likely to get COVID-19 information from

TV (median 5) and online news outlets (median 5), with social media (median 4) and family and friends (median 4) falling in between. As for perceived source credibility, COVID-19 information on traditional media (television, online news outlets, newspapers, and family and friends) was rated by the respondents as more trustworthy (median 5) than that from social media (median 4).

Table 2. Descriptive statistics of key covariates (N=10,063).

Key covariates	Results	Range
Psychological distress, mean (SD)	14.27 (8.00)	0-30
Preventative behavior, median (IQR)	6 (6.33-6.67)	1-7
Perceived information overload, median (IQR)	3 (3-5)	0-6
Perceived susceptibility to COVID-19, median (IQR)	5 (4-7)	1-7
Perceived severity of COVID-19, median (IQR)	6 (5-7)	1-7
Frequency of seeking COVID-19 information from various sources, median (IQR)		
Newspaper	3 (1-5)	0-6
TV	5 (3-6)	0-6
Online news outlet	5 (4-6)	0-6
Social media	4 (3-6)	0-6
Family and friends	4 (3-5)	0-6
Trustworthiness of COVID-19 information from various sources, median (IQR)		
Newspaper	5 (4-6)	1-7
TV	5 (4-6)	1-7
Online news outlet	5 (4-6)	1-7
Social media	4 (4-5)	1-7
Family and friends	5 (4-6)	1-7

Sociodemographic Factors and Perceived Information Overload

Table 3 shows the OLS regression models for perceived information overload and sociodemographic variables. The results showed that older respondents were less likely to experience COVID-19 information overload. Compared with those aged 18 years to 29 years, respondents aged between 40 years and 49 years ($b=-0.11$, 95% CI -0.20 to -0.01 ; $P=.02$), between 50 years and 59 years ($b=-0.13$, 95% CI -0.23 to -0.04 ; $P<.007$), and at least 60 years ($b=-0.21$, 95% CI -0.33 to -0.08 , $P<.001$) expressed lower levels of information overload. Female respondents were more likely to feel information overload ($b=0.20$, 95% CI 0.14 to 0.26 , $P<.001$). In addition, respondents with higher SES were more likely to experience information overload about COVID-19. Compared with those who perceived themselves as lower or lower-middle class, respondents self-perceived as upper or upper-middle class ($b=0.19$, 95% CI 0.09 to 0.30 ; $P<.001$) experienced more COVID-19 information

overload. Those with full-time jobs also tended to perceive higher information overload ($b=0.11$, 95% CI 0.04 to 0.17 ; $P=.003$) than those without full-time work. The relationship between education and perceived information overload was marginally significant, with respondents with at least a college education more likely to be overwhelmed by COVID-19 information. There were no discernible differences in perceived information overload between people living in rural or urban areas and between housing types. In addition, respondents who had pregnant women or older adults (>65 years old) at home ($b=0.16$, 95% CI 0.09 to 0.23 ; $P<.001$) and with children under 12 years in the household ($b=0.35$, 95% CI 0.28 to 0.43 ; $P<.001$) were more likely to perceive COVID-19 information overload. As for RQ1, about which segments of the population experienced greater information overload, our findings revealed that young people, women, full-time workers, those who perceived themselves as upper or upper-middle class, and those with vulnerable populations in the household were more likely to experience COVID-19 information overload.

Table 3. Association between sociodemographic variables and perceived information overload.

Sociodemographic variables	b ^a	95% CI	P value
Age (years)			
18-29	Reference	N/A ^b	N/A
30-39	0.02	–0.07 to 0.11	.70
40-49	–0.11	–0.20 to –0.01	.02
50-59	–0.13	–0.23 to –0.04	.008
≥60	–0.21	–0.33 to –0.08	.002
Sex			
Male	Reference	N/A	N/A
Female	0.20	0.14 to 0.26	<.001
Education			
Secondary or below	Reference	N/A	N/A
College or above	0.08	–0.01 to 0.17	.08
Residential area			
Rural	Reference	N/A	N/A
Urban	0.05	–0.04 to 0.13	.28
Perceived social status			
Lower or lower-middle class	Reference	N/A	N/A
Middle class	–0.02	–0.08 to 0.05	.67
Upper and upper-middle class	0.19	0.09 to 0.30	.001
Employment			
Other	Reference	N/A	N/A
Working full time	0.11	0.04 to 0.17	.003
Housing type			
Other	Reference	N/A	N/A
Privately owned housing	0.01	–0.06 to 0.08	.81
Chronic diseases			
No	Reference	N/A	N/A
Yes	0.18	0.09 to 0.27	<.001
Having pregnant women or older adults (>65 years old) in the household			
No	Reference	N/A	N/A
Yes	0.16	0.09 to 0.23	<.001
Having children aged under 12 years in the household			
No	Reference	N/A	N/A
Yes	0.35	0.28 to 0.43	<.001
Survey location			
Hong Kong	Reference	N/A	N/A
Japan	0.05	–0.06 to 0.17	.37
Singapore	0.57	0.47 to 0.68	<.001
South Korea	0.94	0.84 to 1.05	<.001
Taiwan	–0.41	–0.52 to –0.31	<.001
Thailand	0.66	0.55 to 0.77	<.001

Sociodemographic variables	b ^a	95% CI	P value
Constant	2.64	2.50 to 2.78	<.001

^aUnstandardized coefficient.

^bN/A: not applicable

Information-Seeking Behavior and Perceived Information Overload

We then examined the effects of using different sources for COVID-19 information on perceived information overload after adjusting for perceived source credibility, average time spent on viewing COVID-19 information, and various sociodemographic variables (Table 4). Inconsistent with H1, the results revealed that seeking COVID-19 information from traditional sources was positively associated with perceived information overload. Specifically, respondents who more

frequently sought COVID-19 information from newspapers (b=0.12, 95% CI 0.11 to 0.14; P<.001), television (b=0.07, 95% CI 0.05 to 0.09; P<.001), and from family and friends (b=0.11, 95% CI 0.09 to 0.14; P<.001) were more likely to feel overwhelmed by the information. In contrast, seeking COVID-19 information from online media, including online news outlets (b=0.00, 95% CI -0.03 to 0.03; P=.98) and social media (b=0.02, 95% CI -0.00 to 0.04; P=.10) was not significant associated with perceived information overload, which showed limited support for H2.

Table 4. The association between information seeking behavior and perceived information overload.

Information-seeking behaviors	b ^{a,b}	95% CI	P value
Frequency of seeking COVID-19 information from various sources			
Newspaper	0.12	0.11 to 0.14	<.001
TV	0.07	0.05 to 0.09	<.001
Online news outlet	0.00	-0.03 to 0.03	.98
Social media	0.02	-0.00 to 0.04	.13
Family and friends	0.11	0.09 to 0.14	<.001
Trustworthiness of COVID-19 information from various sources			
Newspaper	-0.00	-0.03 to 0.03	.94
TV	-0.05	-0.08 to -0.01	.01
Online news outlet	0.05	0.01 to 0.08	.01
Social media	0.10	0.07 to 0.13	<.001
Family and friends	0.06	0.03 to 0.10	<.001

^aThe model was adjusted for sociodemographic variables, including age, sex, education, area, perceived social status, employment, housing types, a history of chronic diseases, having pregnant women or older adults (>65 years old) in the household, having children aged under 12 years in the household, and survey locations.

^bUnstandardized coefficient.

The relationships between perceived source credibility and perceived information overload were mixed. Specifically, people who trusted COVID-19 information on television more experienced lower levels of information overload (b=-0.05, 95% CI -0.08 to -0.01; P=.005). In contrast, perceived trustworthiness of COVID-19 information from online news outlets (b=0.05, 95% CI 0.01 to 0.08; P=.005), social media (b=0.10, 95% CI 0.07 to 0.13; P<.001), and family and friends (b=0.06, 95% CI 0.03 to 0.10; P<.001) was positively associated with perceived information overload. The perceived trustworthiness of COVID-19 information on newspapers was not associated with perceived information overload.

Perceived Information Overload and Psychological Distress

Next, we investigated the effect of perceived information overload and information-seeking behavior on psychological distress. Model 1 in Table 5 shows the OLS regression models

for psychological distress and protective behaviors after controlling for sociodemographic variables. As for psychological distress, a positive and significant effect of perceived information overload and psychological distress existed (b=2.18, 95% CI 2.09 to 2.26; P<.001), which lent strong support for H3. As for the role of information-seeking behavior in psychological well-being, the results showed mixed support for H4 and H5. Specifically, more consumption of COVID-19 information from newspapers (b=0.58, 95% CI 0.51 to 0.65; P<.001) and family and friends (b=0.29, 95% CI 0.18 to 0.40; P<.001) resulted in higher levels of psychological distress. In contrast, obtaining COVID-19 information from online news outlets was associated with less distress (b=-0.32, 95% CI -0.44 to -0.21; P<.001). The frequency of getting COVID-19 information from TV and social media was not associated with psychological distress. For perceived source credibility, the perceived trustworthiness of information from newspapers (b=-0.27, 95% CI -0.39 to -0.15; P<.001) and TV (b=-0.26,

95% CI -0.39 to -0.12; $P < .001$) was negatively associated with psychological distress caused by COVID-19. In contrast, the perceived credibility of information obtained from social media

($b = 0.56$, 95% CI 0.43 to 0.68; $P < .001$) and family and friends ($b = 0.24$, 95% CI 0.11 to 0.37; $P < .001$) was positively related to psychological distress.

Table 5. Association between perceived information overload and psychological distress and preventive behaviors.

Variables	Model 1 ^a : psychological distress			Model 2 ^a : preventive behavior		
	b ^b	95% CI	P value	b ^b	95% CI	P value
Perceived information overload	2.18	2.09 to 2.26	<.001	0.01	-0.09 to 0.07	.85
Frequency of seeking COVID-19 information from various sources						
Newspaper	0.58	0.51 to 0.65	<.001	-0.06	-0.13 to 0.00	.06
TV	-0.03	-0.13 to 0.06	.55	0.19	0.10 to 0.29	<.001
Online news outlet	-0.32	-0.44 to -0.21	<.001	0.37	0.26 to 0.47	<.001
Social media	0.05	-0.05 to 0.15	.38	0.18	0.09 to 0.28	<.001
Family and friends	0.29	0.18 to 0.40	<.001	0.01	-0.10 to 0.11	.86
Trustworthiness of COVID-19 information obtained from various sources						
Newspaper	-0.27	-0.39 to -0.15	<.001	0.08	-0.03 to 0.19	.17
TV	-0.26	-0.39 to -0.12	<.001	0.14	0.01 to 0.27	.03
Online news outlet	0.05	-0.09 to 0.19	.55	-0.00	-0.13 to 0.13	.98
Social media	0.56	0.43 to 0.68	<.001	-0.16	-0.27 to -0.03	.01
Family and friends	0.24	0.11 to 0.37	<.001	0.18	0.06 to 0.31	.004
Perceived susceptibility	0.80	0.71 to 0.88	<.001	0.41	0.33 to 0.49	<.001
Perceived severity	0.30	0.20 to 0.39	<.001	0.55	0.46 to 0.63	<.001

^aThe model was adjusted for sociodemographic variables, including age, sex, education, area, perceived social status, employment, housing types, a history of chronic diseases, having pregnant women or older adults (>65 years old) in the household, having children aged under 12 years in the household, and survey locations.

^b Unstandardized coefficient.

Sociodemographic Differences in the Adverse Psychological Effect of Information Overload

Finally, we assessed whether the impacts of COVID-19 information overload on psychological distress varied by different sociodemographic characteristics. A series of 2-way interaction terms between perceived information overload and sociodemographic variables were computed. Table 6 presents the regression results of psychological distress on interactions between perceived information overload and sociodemographic factors. The interaction between information overload and age groups was not significant, suggesting that the effect of information overload was not dependent on the respondents' age—such a finding showed limited support for H6. The results revealed a significant sex difference in the psychological consequences of perceived information overload, with women experiencing more psychological distress in the midst of information overload about COVID-19 ($b = 0.24$, 95% CI 0.10 to 0.39; $P = .001$).

As for SES, there was a positive interaction between COVID-19 information overload and an urban residence ($b = 0.23$, 95% CI

0.04 to 0.41; $P = .02$), suggesting that the detrimental effect of COVID-19 information overload on psychological distress was more salient for urban residents than their rural counterparts. The interaction between information overload and middle-class status was also significant ($b = 0.18$, 95% CI -0.02 to 0.33; $P = .03$), suggesting that self-perceived middle-class respondents were more likely to experience psychological distress when faced with COVID-19 information overload than respondents self-perceived as being lower or lower-middle class. Similarly, full-time employees were likely to experience higher levels of psychological distress when perceiving information overload about COVID-19 ($b = 0.24$, 95% CI 0.09 to 0.39; $P = .002$). Also, the association between perceived information overload and psychological distress was more salient among respondents living in privately owned housing than their counterparts residing in other types of accommodation ($b = 0.20$, 95% CI 0.05 to 0.35; $P = .01$). Inconsistent with H7, such results suggested that those with higher SES were more likely to develop psychological distress when experiencing COVID-19 information overload.

Table 6. Associations between perceived information overload and psychological distress among Asian populations with different sociodemographic backgrounds.

Models and variables	b ^a	95% CI	P value
Model 1^b: perceived IO^c × age			
Perceived IO	2.08	1.93 to 2.24	<.001
Age: 18-29 years	Reference	N/A ^d	N/A
Age: 30-39 years	-0.67	-1.48 to 0.13	.10
Age: 40-49 years	-1.26	-2.07 to -0.46	.002
Age: 50-59 years	-1.35	-2.18 to -0.55	.001
Age: ≥60 years	-3.23	-4.28 to -2.18	<.001
Perceived IO × 30-39 years	0.18	-0.03 to 0.39	.10
Perceived IO × 40-49 years	0.18	-0.03 to 0.40	.09
Perceived IO × 50-59 years	-0.01	-0.23 to 0.21	.94
Perceived IO × ≥60 years	0.13	-0.16 to 0.41	.39
Model 2^b: perceived IO × sex			
Perceived IO	2.07	1.96 to 2.17	<.001
Sex: male	Reference	N/A	N/A
Sex: female	-1.23	-1.77 to -0.68	<.001
Perceived IO × female	0.24	0.10 to 0.39	.001
Model 3^b: perceived IO × education			
Perceived IO	2.16	1.98 to 2.34	<.001
Education: secondary or below	Reference	N/A	N/A
Education: college or above	-0.05	-0.75 to 0.65	.90
Perceived IO × college or above	0.02	-0.17 to 0.21	.83
Model 4^b: perceived IO × area			
Perceived IO	2.00	1.82 to 2.17	<.001
Residential area: rural	Reference	N/A	N/A
Residential area: urban	-0.53	-1.22 to 0.16	.13
Perceived IO × urban	0.23	0.04 to 0.41	.02
Model 5^b: perceived IO × perceived status			
Perceived IO	2.10	1.98 to 2.23	<.001
Perceived status: lower or lower-middle class	Reference	N/A	N/A
Perceived status: middle class	-0.79	-1.37 to -0.20	.008
Perceived status: upper and upper-middle class	0.62	-0.27 to 1.52	.17
Perceived IO × middle class	0.18	0.02 to 0.33	.03
Perceived IO × upper or upper-middle class	-0.07	-0.31 to 0.16	.53
Model 6^b: perceived IO × employment status			
Perceived IO	2.03	1.91 to 2.16	<.001
Employment status: non-full time work	Reference	N/A	N/A
Employment status: working full time	-0.76	-1.33 to -0.20	.008
Perceived IO × working fulltime	0.24	0.09 to 0.39	.002
Model 7^b: perceived IO × housing types			
Perceived IO	2.06	1.93 to 2.18	<.001

Models and variables	b ^a	95% CI	P value
Housing type: non-privately owned housing	Reference	N/A	N/A
Housing type: privately owned housing	-0.77	-1.34 to -0.21	.007
Perceived IO × privately-owned housing	0.20	0.05 to 0.35	.01

^aUnstandardized coefficient.

^bAll the models adjusted for sociodemographic variables, including age, sex, education, area, perceived social status, employment, housing types, a history of chronic diseases, having pregnant women or older adults (>65 years old) in the household, having children aged under 12 years in the household, and survey locations. The model also controlled for perceived susceptibility to and perceived severity of COVID-19.

^cIO: information overload.

^dN/A: not applicable.

Perceived Information Overload and Protective Behavior

Model 2 in Table 5 shows that the association between perceived information overload and adoption of protective behaviors was not significant after adjusting for sociodemographic variables and perceived susceptibility and severity of the pandemic ($b=0.00$, 95% CI -0.01 to 0.01 ; $P=.88$). Such results did not support H8. More frequent exposure to COVID-19 information on TV ($b=0.33$, 95% CI 0.01 to 0.04 ; $P<.001$), online news outlets ($b=0.06$, 95% CI 0.04 to 0.08 ; $P<.001$), and social media ($b=0.03$, 95% CI 0.01 to 0.04 ; $P=.001$) could promote the adoption of protective behaviors against COVID-19. Also, respondents who perceived higher trustworthiness of COVID-19 information obtained from TV ($b=0.03$, 95% CI 0.01 to 0.05 ; $P=.007$) and family and friends ($b=0.03$, 95% CI 0.01 to 0.05 ; $P=.008$) were more likely to take protective measures. In contrast, there was a negative association between perceived trustworthiness of COVID-19 information on social media and engagement in various protective behaviors ($b=-0.02$, 95% CI -0.04 to 0.00 ; $P=.04$).

Discussion

Principal Findings

This study is among the first to investigate the antecedents and consequences of information overload about COVID-19 among Asian populations. Using data from a cross-sectional survey of 10,063 residents of 6 jurisdictions in East and Southeast Asia, our study showed a high level of perceived information overload during the pandemic. Regression results further revealed that young people, women, people with a higher SES (ie, full-time workers, self-perception as being upper or upper-middle class), and those with vulnerable populations in the household were more likely to experience COVID-19 information overload. As for the behavioral consequence of information overload, the results showed no significant relationship between perceived information overload and protective behaviors during the pandemic. Consistent with previous studies, we found a positive relationship between perceived information overload and psychological distress. Notably, the association between perceived information overload and psychological distress was more substantial among women and people with a higher SES (urban residents, self-perceived as middle class, full-time workers, and people living in privately owned housing). The findings of this study contribute to a better understanding of the

level and correlates of information overload during the COVID-19 pandemic and help identify subpopulations that are particularly susceptible to information overload and its potential downstream consequences. We discuss the main findings in the following sections.

Comparison With Prior Work

Although the occurrence of information overload has been documented in other disease outbreaks, the level and consequences of information overload during the current global pandemic are unparalleled. On the one hand, the unprecedented scale and impacts of COVID-19 on public health and individual lives have led to intensive media coverage. Other epidemics such as Zika, Ebola, Middle East Respiratory Syndrome (MERS), and H1N1 (swine flu) have caused great damage in daily life. Still, the level of panic caused by COVID-19 is much more severe and has resulted in a large volume of news attention to COVID-19. Also, the high scientific uncertainty and rapidly evolving settings of COVID-19 create opportunities for generation of ambiguous, inaccurate, and conflicting information, which may amplify information overload. On the other hand, the media environment and spreading of information have significantly changed over the past several decades. When another deadly viral disease—Severe Acute Respiratory Syndrome (SARS)—broke out in 2003, none of the major social media outlets was present. In pandemics before the social media era, a multilayered process involving careful review by experts, editing by journals, and releasing essential information under embargo to journalists who seek third-party comments would be performed before releasing scientific knowledge to the public [71]. In contrast, social media platforms can host an enormous amount of user-generated content. During the COVID-19 pandemic, there has been a rush to publish unverified information or even deliberately spread misinformation on various social media platforms, which is likely to cause panic and noncompliance to preventive behaviors.

As for the first research question about which segments of the population perceived higher levels of information overload during the COVID-19 pandemic, our results showed that younger people (18-39 years old), women, and respondents with higher SES (having a college education or above, having a full-time job, and self-perceived as upper or upper-middle class) expressed higher levels of perceived information overload about COVID-19. The finding that women tended to experience more information overload about COVID-19 than men is consistent with findings of previous studies [40-42]. Globally, 70% of the

workers in the health and social sectors are women [72]. Women provide the majority of health services and are more likely to take the caregiving lead in the family. Women's caregiving roles and higher risk exposure during the pandemic have prompted them to pay more attention to COVID-19 information and take COVID-19 more seriously than men [73,74].

Though not consistent with previous studies [31], the finding that older people experienced lower levels of information overload about COVID-19 may be justified by the following reasons. First, increased age is often associated with decreased motivation for health-related information seeking, especially from online media, as older adults typically have lower levels of internet literacy and may experience more difficulties navigating websites [75]. Moreover, older people may be more experienced with information content, even on online platforms. In contrast to younger online users, who mainly browse entertaining and social network sites, older adults are more likely to use the internet for information purposes [76]. Thus, the older generation is less likely to feel cognitively overloaded as they may not seek COVID-19 information as frequently as young people, and they are more experienced in dealing with information [45].

Although several studies have suggested that people with higher SES may experience less information overload [31,36-39], our results showed a positive relationship between education, employment, and perceived social status with perceived information overload about COVID-19. Such unexpected findings may be due to the widening digital divide between socioeconomic groups during the COVID-19 pandemic. The divides in internet coverage and quality of service have exposed people to different levels of life disruptions caused by the pandemic. Although those with adequate internet infrastructure and capacity can work from home, engage in online teaching and learning, and order food and groceries online, people with limited digital access and skills can hardly overcome the economic hardship and life inconvenience caused by the COVID-19 pandemic. More importantly, the costs for timely information for digital "have-nots" tend to be high, as they may need to wait for information to arrive informally rather than obtaining the information online in real time. Given that people with lower SES are more likely to be digitally disadvantaged, they may be involved in fewer information-seeking behaviors and, consequently, less likely to feel overwhelmed by COVID-19 information.

Moreover, the socioeconomic disparities in perceived information overload may gradually emerge as the pandemic unfolds. A study of South Korean residents found no sociodemographic differences in perceived information overload about COVID-19 during the early stage of the pandemic [25]. It may be because their study was conducted at the peak of media coverage of COVID-19 in South Korea (mid-March 2020), and there may be a ceiling effect of perceived information overload when most people were highly attentive to the COVID-19 information. Moreover, the data collection was confined to Seoul residents, and the sample overrepresented young and highly educated people, which limited the socioeconomic differences in the perception of information overload. In contrast, our study was carried out in May 2020

when the 6 jurisdictions surveyed were in their fifth month of the pandemic. Our findings suggested that the disparities in the perception of information overload may gradually emerge as the pandemic unfolds. Also, our study included representative samples in the 6 jurisdictions in East and Southeast Asia, which reduced sampling bias and increased the generalizability of the findings to the target population. Future studies may further examine whether the current results are consistent across different contexts and stages of the pandemic.

In addition, respondents with vulnerable family members were more sensitive to the threats from COVID-19 and, accordingly, were more attentive to relevant information. Our results support that argument by revealing that respondents with vulnerable significant others at home (eg, pregnant women, young children, or people over 65 years old) were more likely to feel overwhelmed by COVID-19 information. However, inconsistent with previous studies showing that people who perceived themselves to have less-than-excellent health were more susceptible to information overload [77], our results indicated an insignificant relationship between chronic illness records and perceived information overload about COVID-19. More studies are needed to explore the association between pre-existing health conditions and perceived information overload during the COVID-19 pandemic.

As for the effect of information-seeking behavior on perceived information overload (second research question), we found that the use of traditional media (eg, newspaper, television) for seeking COVID-19 information was positively associated with perceived information overload. In contrast, the effect of getting COVID-19 information from online news outlets was not significant. Given that the internet and online media contribute to the over-proliferation of health information available to the lay public [51], such results seem unreasonable. A possible explanation is that, while offline seekers have little or no control over what is aired but receive information passively, online information seekers can adjust the search terms anytime to suit their search needs, thereby getting tailored information. Compared with traditional media, internet media are active channels that require greater cognitive effort and are often used by highly motivated and engaged individuals [78]. Moreover, online users can easily search other websites for clarification or fact-checking whenever they find the information received from a single online source is unreliable [17]. Thus, the motivation to authenticate earlier information obtained may prohibit online seekers from being overwhelmed with the information received. Future studies should take the route of information acquisition (eg, active search vs passive exposure) into account.

Our results underscore the critical role of trust in information sources in crisis management. Interestingly, the perceived reliability of COVID-19 information from various media channels exerted differential effects on perceived information overload in our study. Perceived trustworthiness of COVID-19 information from television was negatively associated with perceived information overload. It was consistent with prior research that higher trust in information sources predicted less distress by information and benefited crisis management [79]. However, respondents were more likely to experience

information overload when they perceived COVID-19 information from online media (ie, online news outlets, social media) as more trustworthy. It may be because misinformation, disinformation, and rumors about COVID-19 were more prevalent on online media than other forms of media [44], which may engender greater confusion and require more cognitive effort to process such inaccurate and inconsistent information for those who believe in the information. As mentioned earlier, the frequency of seeking COVID-19 information online was not associated with perceived information overload. Thus, the level of perceived information overload seems more attributable to the perceived credibility of information on the internet rather than the online information-seeking behavior.

As for the third research question about the psychological and behavioral consequences of information overload, we found that an overabundance of COVID-19 information can harm mental well-being by increasing the likelihood of experiencing posttraumatic stress disorder. Such a finding was consistent with previous studies about the negative psychological impact of information overload [17]. During an unprecedented global pandemic like COVID-19, people are understandably attentive to health information and, by doing so, try to reduce the negative affect caused by the previously unknown and unpredictable disease. However, when information flow exceeds one's capacity to process, the perceived information overload may cause a detrimental effect on psychological well-being. Moreover, deteriorated mental health caused by information overload may lead to information avoidance when people deliberately avoid seeking health information to lessen the cognitive burden and negative affect [76]. Information avoidance is detrimental for public health because the acquisition of health information helps individuals make informed medical decisions and engage in preventive behaviors [80]. Given the unintended consequences of information overload, our study highlighted the importance of keeping a balance between information transparency and information overload.

However, our results did not show a significant effect of perceived information overload on protective behaviors during the COVID-19 pandemic, which was consistent with a study conducted during the COVID-19 pandemic in Korea [25]. Despite no direct effect, information overload may indirectly reduce compliance via increasing heuristic processing and decreasing systematic processing [25]. Compared with systematic processing of information, judgments based on heuristic processing tend to be less stable and weakly tied to subsequent behaviors [81], which may discourage one's willingness to adopt protective measures. When the risk perception of COVID-19 decreases and the anxiety associated with this pandemic fade over time, the detrimental effect of perceived information overload on protective behaviors may emerge. Future studies are warranted to investigate the behavioral consequences of information overload in different time frames.

The fourth research question focused on who was more vulnerable to the psychological and behavioral consequences of information overload. Since there was no significant effect of perceived information overload on protective behaviors, we did not further examine whether such an effect may be different

across various sociodemographic characteristics. As for psychological distress, we found that women and people with a higher SES were more vulnerable to the adverse effects of perceived information overload on psychological well-being. Female respondents were more likely to experience psychological distress when exposed to information overload, possibly because women are more likely than men to perceive the pandemic as a severe health problem and to agree and comply with restraining measures [53]. Besides, we speculated that people with a higher SES might be less psychologically resilient in the face of the COVID-19 pandemic and thus perceived more significant information overload. Previous studies have suggested that people with a higher SES were more concerned and worried about the pandemic [82] and may have been experiencing a more drastic relative change in their lifestyles, leading to a sharper decline in their psychological well-being.

Meanwhile, individuals with a lower SES tend to experience more everyday stressors even when there is no global pandemic. Thus, they may not undergo a significant change in their subjective well-being when reading pessimistic news about COVID-19. Besides, people with higher education care more about the quality of information and may feel frustrated if they cannot find trustworthy sources during the pandemic. Given the ambiguous information and inconsistent guidance about COVID-19, especially during the early stage of the outbreak, it is hard to identify valuable and reliable information even for people with a higher SES who are believed to have more access to health information and a higher ability to process such information.

Since information overload can harm mental well-being and potentially reduce compliance during the pandemic, it is urgent to manage the overload of information that exceeds people's cognitive ability to process. From the information provider side, government and media should disseminate evidence-based and transparent information swiftly and widely among the public. Social media technology companies must constantly review content shared on their platforms and closely verify the reliability of information related to the pandemic. Since there are stratified mechanisms and consequences of information overload as shown in this study, information policies and management should be accordingly "stratified" as well. It is essential to develop efficient health communication strategies targeting people with different sociodemographic characteristics. Certain subgroups may be more frustrated with the uncertainty caused by the pandemic and eagerly look for sources to fill their information needs. It is necessary to formulate different information dissemination strategies in terms of information content and channels for different groups. It is also essential to establish interventions to help people vulnerable to information overload to cope with information anxiety and mental health concerns.

On the receiver side, individuals should carry out an "information diet" by controlling the extent and type of information they consume. Researchers have suggested that people visit authentic and official websites for COVID-19 information and try to verify suspicious news on one of the many fact-checking websites dedicated to debunking myths

[8,83]. For example, the World Health Organization's (WHO) risk communication team launched a new information platform known as the WHO Information Network for Epidemics, which actively reports details related to COVID-19. A team of WHO "myth busters" works with search and media companies like Facebook, Google, Pinterest, Tencent, Twitter, TikTok, YouTube, and others to counter the spread of rumors. Also, enhancing information literacy skills, particularly health literacy skills, can facilitate the critical evaluation of health information and help people make appropriate health-related decisions. It is argued that information and health literacy is vital to reducing information overload and its consequences [84,85]. Thus, extensive training and guidelines for improving information and health literacy should be promoted to the public, particularly regarding crisis management.

Limitations

Despite the significant findings, this study is not without limitations. First, the data were obtained from a cross-sectional survey, and it is hard to ascertain the causal relationship between variables. For example, our results showed a negative effect of perceived information overload on psychological well-being. However, trait anxiety was significantly associated with higher levels of perceived information overload [24]. Thus, longitudinal studies are needed to ascertain the psychological consequences of information overload. Second, only one item was used to measure perceived information overload, which presents a significant limitation of this study. We recognized that information overload is a multidimensional construct and a single proxy item limits our ability to make clear-cut

generalizations from our findings. Future studies may adopt validated scales regarding COVID-19 information overload (eg, Corona Information Overload Scale [65]). Third, this study has limited information on knowledge and attitudes related to COVID-19, though we have adjusted for perceived susceptibility and perceived severity of COVID-19 in the analysis. Future studies may further examine how knowledge and attitudes may be associated with information overload.

Conclusion

Perceptions of information overload are prevalent during the COVID-19 pandemic and have caused significant psychological and behavioral consequences. This study is among the first to examine how the antecedents and consequences of perceived information overload vary between different sociodemographic groups among the Asian population. A cross-sectional survey with representative data of 10,063 residents in 6 jurisdictions in Asia was conducted in May 2020. Regression results confirmed a positive relationship between perceived information overload and psychological distress. Also, people with a higher SES were more exposed to information overload about COVID-19 and suffered more psychological distress because of perceived information overload. Our findings suggested that the provision of trustworthy information and reduction of the perceived information overload can significantly ameliorate psychological distress during the pandemic. Effective policies and interventions should be promoted to target higher-SES populations who are more susceptible to the occurrence and adverse psychological influence of perceived information overload.

Acknowledgments

This study was supported by a Knowledge Transfer Earmarked Fund from Hong Kong University Grants Commission (6354048). The funder had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Tests of Ordinal Least Squares (OLS) assumptions.

[DOCX File, 460 KB - [jmir_v24i3e31088_app1.docx](#)]

References

1. Mandhana N. Asia Suffers Outbreaks Where Covid-19 Had Seemed Beaten. Wall Street Journal. 2021 May 19. URL: <https://www.wsj.com/articles/as-covid-19-variants-spread-and-vaccines-dont-asia-suffers-outbreaks-where-disease-had-seemed-beaten-11621419961> [accessed 2022-02-25]
2. Coronavirus updates. Worldometer. URL: https://www.worldometers.info/coronavirus/?utm_campaign=homeAdUOA?Si [accessed 2021-05-31]
3. Jin L, Chen X, Lin F, Zou Y, Gao H. Does education matter for psychological recovery amidst the COVID-19 pandemic? Evidence from a panel survey in Hubei, China. *Anxiety Stress Coping* 2022 Jan;35(1):101-110. [doi: [10.1080/10615806.2021.1978431](https://doi.org/10.1080/10615806.2021.1978431)] [Medline: [34550829](#)]
4. Tso IF, Park S. Alarming levels of psychiatric symptoms and the role of loneliness during the COVID-19 epidemic: A case study of Hong Kong. *Psychiatry Res* 2020 Nov;293:113423 [FREE Full text] [doi: [10.1016/j.psychres.2020.113423](https://doi.org/10.1016/j.psychres.2020.113423)] [Medline: [32871487](#)]

5. Lee H, Dean D, Baxter T, Griffith T, Park S. Deterioration of mental health despite successful control of the COVID-19 pandemic in South Korea. *Psychiatry Res* 2021 Jan;295:113570 [FREE Full text] [doi: [10.1016/j.psychres.2020.113570](https://doi.org/10.1016/j.psychres.2020.113570)] [Medline: [33234326](https://pubmed.ncbi.nlm.nih.gov/33234326/)]
6. Li D, Ko N, Chen Y, Wang P, Chang Y, Yen C, et al. COVID-19-related factors associated with sleep disturbance and suicidal thoughts among the Taiwanese public: a Facebook survey. *Int J Environ Res Public Health* 2020 Jun 22;17(12):4479 [FREE Full text] [doi: [10.3390/ijerph17124479](https://doi.org/10.3390/ijerph17124479)] [Medline: [32580433](https://pubmed.ncbi.nlm.nih.gov/32580433/)]
7. Novel coronavirus (2019-nCov): situation report-13. World Health Organization. 2020 Feb 02. URL: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf> [accessed 2022-02-25]
8. Rathore F, Farooq F. Information overload and infodemic in the COVID-19 pandemic. *J Pak Med Assoc* 2020;1. [doi: [10.5455/jpma.38](https://doi.org/10.5455/jpma.38)]
9. Taha S, Matheson K, Cronin T, Anisman H. Intolerance of uncertainty, appraisals, coping, and anxiety: the case of the 2009 H1N1 pandemic. *Br J Health Psychol* 2014 Sep 09;19(3):592-605. [doi: [10.1111/bjhp.12058](https://doi.org/10.1111/bjhp.12058)] [Medline: [23834735](https://pubmed.ncbi.nlm.nih.gov/23834735/)]
10. Fischhoff B, Wong-Parodi G, Garfin DR, Holman EA, Silver RC. Public understanding of Ebola risks: mastering an unfamiliar threat. *Risk Anal* 2018 Jan 08;38(1):71-83. [doi: [10.1111/risa.12794](https://doi.org/10.1111/risa.12794)] [Medline: [28597480](https://pubmed.ncbi.nlm.nih.gov/28597480/)]
11. Nagler RH, Vogel RI, Gollust SE, Rothman AJ, Fowler EF, Yzer MC. Public perceptions of conflicting information surrounding COVID-19: Results from a nationally representative survey of U.S. adults. *PLoS One* 2020 Oct 21;15(10):e0240776 [FREE Full text] [doi: [10.1371/journal.pone.0240776](https://doi.org/10.1371/journal.pone.0240776)] [Medline: [33085719](https://pubmed.ncbi.nlm.nih.gov/33085719/)]
12. 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](https://doi.org/10.1037/hea0000875)] [Medline: [32202824](https://pubmed.ncbi.nlm.nih.gov/32202824/)]
13. Sweeny K, Melnyk D, Miller W, Shepperd JA. Information avoidance: who, what, when, and why. *Review of General Psychology* 2010 Dec 01;14(4):340-353. [doi: [10.1037/a0021288](https://doi.org/10.1037/a0021288)]
14. Jensen JD, Carcioppolo N, King AJ, Scherr CL, Jones CL, Niederdeppe J. The cancer information overload (CIO) scale: establishing predictive and discriminant validity. *Patient Educ Couns* 2014 Jan;94(1):90-96. [doi: [10.1016/j.pec.2013.09.016](https://doi.org/10.1016/j.pec.2013.09.016)] [Medline: [24268921](https://pubmed.ncbi.nlm.nih.gov/24268921/)]
15. Niederdeppe J, Fowler E, Goldstein K, Pribble J. Does local television news coverage cultivate fatalistic beliefs about cancer prevention? *J Commun* 2010 Jun 01;60(2):230-253 [FREE Full text] [doi: [10.1111/j.1460-2466.2009.01474.x](https://doi.org/10.1111/j.1460-2466.2009.01474.x)] [Medline: [20563221](https://pubmed.ncbi.nlm.nih.gov/20563221/)]
16. Liu C, Kuo K. Does information overload prevent chronic patients from reading self-management educational materials? *Int J Med Inform* 2016 May;89:1-8. [doi: [10.1016/j.ijmedinf.2016.01.012](https://doi.org/10.1016/j.ijmedinf.2016.01.012)] [Medline: [26980353](https://pubmed.ncbi.nlm.nih.gov/26980353/)]
17. Swar B, Hameed T, Reyachav I. Information overload, psychological ill-being, and behavioral intention to continue online healthcare information search. *Computers in Human Behavior* 2017 May;70:416-425. [doi: [10.1016/j.chb.2016.12.068](https://doi.org/10.1016/j.chb.2016.12.068)]
18. Crook B, Stephens KK, Pastorek AE, Mackert M, Donovan EE. Sharing health information and influencing behavioral intentions: the role of health literacy, information overload, and the internet in the diffusion of healthy heart information. *Health Commun* 2016;31(1):60-71. [doi: [10.1080/10410236.2014.936336](https://doi.org/10.1080/10410236.2014.936336)] [Medline: [25668744](https://pubmed.ncbi.nlm.nih.gov/25668744/)]
19. Niederdeppe J, Lee T, Robbins R, Kim HK, Kresovich A, Kirshenblat D, et al. Content and effects of news stories about uncertain cancer causes and preventive behaviors. *Health Commun* 2014 Jun 21;29(4):332-346. [doi: [10.1080/10410236.2012.755603](https://doi.org/10.1080/10410236.2012.755603)] [Medline: [23790111](https://pubmed.ncbi.nlm.nih.gov/23790111/)]
20. No authors listed. The COVID-19 infodemic. *The Lancet Infectious Diseases* 2020 Aug;20(8):875 [FREE Full text] [doi: [10.1016/S1473-3099\(20\)30565-X](https://doi.org/10.1016/S1473-3099(20)30565-X)]
21. Mohammed M, Sha'aban A, Jatau AI, Yunusa I, Isa AM, Wada AS, et al. Assessment of COVID-19 information overload among the general public. *J Racial Ethn Health Disparities* 2022 Feb 19;9(1):184-192. [doi: [10.1007/s40615-020-00942-0](https://doi.org/10.1007/s40615-020-00942-0)] [Medline: [33469869](https://pubmed.ncbi.nlm.nih.gov/33469869/)]
22. Liu H, Liu W, Yoganathan V, Osburg V. COVID-19 information overload and generation Z's social media discontinuance intention during the pandemic lockdown. *Technol Forecast Soc Change* 2021 May;166:120600 [FREE Full text] [doi: [10.1016/j.techfore.2021.120600](https://doi.org/10.1016/j.techfore.2021.120600)] [Medline: [34876758](https://pubmed.ncbi.nlm.nih.gov/34876758/)]
23. Bala R, Srivastava A, Ningthoujam GD, Potsangbam T, Oinam A, Anal CL. An observational study in Manipur State, India on preventive behavior influenced by social media during the COVID-19 pandemic mediated by cyberchondria and information overload. *J Prev Med Public Health* 2021 Jan;54(1):22-30 [FREE Full text] [doi: [10.3961/jpmph.20.465](https://doi.org/10.3961/jpmph.20.465)] [Medline: [33618496](https://pubmed.ncbi.nlm.nih.gov/33618496/)]
24. Fan J, Smith AP. Information overload, wellbeing and COVID-19: a survey in China. *Behav Sci (Basel)* 2021 Apr 27;11(5):62 [FREE Full text] [doi: [10.3390/bs11050062](https://doi.org/10.3390/bs11050062)] [Medline: [33925611](https://pubmed.ncbi.nlm.nih.gov/33925611/)]
25. Hong H, Kim HJ. Antecedents and consequences of information overload in the COVID-19 pandemic. *Int J Environ Res Public Health* 2020 Dec 12;17(24):9305 [FREE Full text] [doi: [10.3390/ijerph17249305](https://doi.org/10.3390/ijerph17249305)] [Medline: [33322715](https://pubmed.ncbi.nlm.nih.gov/33322715/)]
26. Sokolov M. The pandemic infodemic: how social media helps (and hurts) during the coronavirus outbreak. *The Drum*. 2020 Mar 03. URL: <https://www.thedrum.com/opinion/2020/03/03/the-pandemic-infodemic-how-social-media-helps-and-hurts-during-the-coronavirus> [accessed 2022-02-25]

27. Yang Y, Liu K, Li S, Shu M. Social media activities, emotion regulation strategies, and their interactions on people's mental health in COVID-19 pandemic. *Int J Environ Res Public Health* 2020 Dec 01;17(23):8931 [FREE Full text] [doi: [10.3390/ijerph17238931](https://doi.org/10.3390/ijerph17238931)] [Medline: [33271779](https://pubmed.ncbi.nlm.nih.gov/33271779/)]
28. Misra S, Stokols D. Psychological and health outcomes of perceived information overload. *Environment and Behavior* 2011 Apr 17;44(6):737-759. [doi: [10.1177/0013916511404408](https://doi.org/10.1177/0013916511404408)]
29. Khaleel I, Wimmer BC, Peterson GM, Zaidi STR, Roehrer E, Cummings E, et al. Health information overload among health consumers: A scoping review. *Patient Educ Couns* 2020 Jan;103(1):15-32. [doi: [10.1016/j.pec.2019.08.008](https://doi.org/10.1016/j.pec.2019.08.008)] [Medline: [31451363](https://pubmed.ncbi.nlm.nih.gov/31451363/)]
30. Bawden D, Robinson L. The dark side of information: overload, anxiety and other paradoxes and pathologies. *Journal of Information Science* 2008 Nov 21;35(2):180-191. [doi: [10.1177/0165551508095781](https://doi.org/10.1177/0165551508095781)]
31. Chan YM, Huang H. Weight management information overload challenges in 2007 HINTS: socioeconomic, health status and behaviors correlates. *Journal of Consumer Health On the Internet* 2013 Apr;17(2):151-167. [doi: [10.1080/15398285.2013.780540](https://doi.org/10.1080/15398285.2013.780540)]
32. Akin LK. Information overload: A multi-disciplinary explication and citation ranking within three selected disciplines: Library studies, psychology/psychiatry, and consumer science: 1960-1996. Texas Woman's University. 1997 May. URL: <https://twu-ir.tdl.org/handle/11274/10267> [accessed 2022-02-25]
33. Ji Q, Ha L, Sypher U. The role of news media use and demographic characteristics in the possibility of information overload prediction. *International Journal of Communication* 2014;8-24 [FREE Full text]
34. Lagoe C, Atkin D. Health anxiety in the digital age: An exploration of psychological determinants of online health information seeking. *Computers in Human Behavior* 2015 Nov;52:484-491. [doi: [10.1016/j.chb.2015.06.003](https://doi.org/10.1016/j.chb.2015.06.003)]
35. York C. Overloaded by the news: effects of news exposure and enjoyment on reporting information overload. *Communication Research Reports* 2013 Oct;30(4):282-292. [doi: [10.1080/08824096.2013.836628](https://doi.org/10.1080/08824096.2013.836628)]
36. Chae J, Lee C, Jensen JD. Correlates of cancer information overload: focusing on individual ability and motivation. *Health Commun* 2016 Oct 29;31(5):626-634. [doi: [10.1080/10410236.2014.986026](https://doi.org/10.1080/10410236.2014.986026)] [Medline: [26512760](https://pubmed.ncbi.nlm.nih.gov/26512760/)]
37. Obamiro K, Lee K. Information overload in patients with atrial fibrillation: Can the cancer information overload (CIO) scale be used? *Patient Educ Couns* 2019 Mar;102(3):550-554. [doi: [10.1016/j.pec.2018.10.005](https://doi.org/10.1016/j.pec.2018.10.005)] [Medline: [30314829](https://pubmed.ncbi.nlm.nih.gov/30314829/)]
38. Kim K, Lustria ML, Burke D, Kwon N. Predictors of cancer information overload: findings from a national survey. *Information Research* 2007;12(4):1 [FREE Full text]
39. Vollmer Dahlke D, Fair K, Hong YA, Beaudoin CE, Pulczynski J, Ory MG. Apps seeking theories: results of a study on the use of health behavior change theories in cancer survivorship mobile apps. *JMIR Mhealth Uhealth* 2015 Mar 27;3(1):e31 [FREE Full text] [doi: [10.2196/mhealth.3861](https://doi.org/10.2196/mhealth.3861)] [Medline: [25830810](https://pubmed.ncbi.nlm.nih.gov/25830810/)]
40. Jensen JD, Liu M, Carcioppolo N, John KK, Krakow M, Sun Y. Health information seeking and scanning among US adults aged 50-75 years: Testing a key postulate of the information overload model. *Health Informatics J* 2017 Jun 23;23(2):96-108 [FREE Full text] [doi: [10.1177/1460458215627290](https://doi.org/10.1177/1460458215627290)] [Medline: [26905079](https://pubmed.ncbi.nlm.nih.gov/26905079/)]
41. Gurmankin AD, Viswanath V. The impact of health information overload on attention to information and cancer risk reduction behavior. URL: <https://slideplayer.com/slide/6292688/> [accessed 2022-02-25]
42. Chao M, Xue D, Liu T, Yang H, Hall BJ. Media use and acute psychological outcomes during COVID-19 outbreak in China. *J Anxiety Disord* 2020 Aug;74:102248 [FREE Full text] [doi: [10.1016/j.janxdis.2020.102248](https://doi.org/10.1016/j.janxdis.2020.102248)] [Medline: [32505918](https://pubmed.ncbi.nlm.nih.gov/32505918/)]
43. Merchant RM, Lurie N. Social media and emergency preparedness in response to novel coronavirus. *JAMA* 2020 May 26;323(20):2011-2012. [doi: [10.1001/jama.2020.4469](https://doi.org/10.1001/jama.2020.4469)] [Medline: [32202611](https://pubmed.ncbi.nlm.nih.gov/32202611/)]
44. Bastani P, Bahrami M. COVID-19 related misinformation on social media: a qualitative study from Iran. *J Med Internet Res* 2020 Apr 05;5 [FREE Full text] [doi: [10.2196/18932](https://doi.org/10.2196/18932)] [Medline: [32250961](https://pubmed.ncbi.nlm.nih.gov/32250961/)]
45. Eppler MJ, Mengis J. Side-effects of the e-society: The causes of information overload and possible countermeasures. 2004 Presented at: IADIS International Conference e-Society; July 7, 2004; Ávila, Spain.
46. Holmes BJ. Communicating about emerging infectious disease: The importance of research. *Health, Risk & Society* 2008 Aug;10(4):349-360. [doi: [10.1080/13698570802166431](https://doi.org/10.1080/13698570802166431)]
47. Tausczik Y, Faasse K, Pennebaker JW, Petrie KJ. Public anxiety and information seeking following the H1N1 outbreak: blogs, newspaper articles, and Wikipedia visits. *Health Commun* 2012 Aug 09;27(2):179-185. [doi: [10.1080/10410236.2011.571759](https://doi.org/10.1080/10410236.2011.571759)] [Medline: [21827326](https://pubmed.ncbi.nlm.nih.gov/21827326/)]
48. Chaiken S, Ledgerwood A. A theory of heuristic and systematic information processing. In: Van Lange PAM, Kruglanski AW, Higgins ET, editors. *Handbook of theories of social psychology*. Thousand Oaks, CA: Sage Publications Ltd; 2012:246-266.
49. Wheaton MG, Abramowitz JS, Berman NC, Fabricant LE, Olatunji BO. Psychological predictors of anxiety in response to the H1N1 (Swine Flu) pandemic. *Cogn Ther Res* 2011 Feb 22;36(3):210-218. [doi: [10.1007/s10608-011-9353-3](https://doi.org/10.1007/s10608-011-9353-3)]
50. Thompson RR, Garfin DR, Holman EA, Silver RC. Distress, worry, and functioning following a global health crisis: a national study of Americans' responses to Ebola. *Clinical Psychological Science* 2017 Apr 26;5(3):513-521. [doi: [10.1177/2167702617692030](https://doi.org/10.1177/2167702617692030)]

51. 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):1 [FREE Full text] [doi: [10.3390/ijerph17051729](https://doi.org/10.3390/ijerph17051729)] [Medline: [32155789](https://pubmed.ncbi.nlm.nih.gov/32155789/)]
52. 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 Apr 16;15(4):e0231924 [FREE Full text] [doi: [10.1371/journal.pone.0231924](https://doi.org/10.1371/journal.pone.0231924)] [Medline: [32298385](https://pubmed.ncbi.nlm.nih.gov/32298385/)]
53. Ko N, Lu W, Chen Y, Li D, Wang P, Hsu S, et al. COVID-19-related information sources and psychological well-being: An online survey study in Taiwan. *Brain Behav Immun* 2020 Jul;87:153-154 [FREE Full text] [doi: [10.1016/j.bbi.2020.05.019](https://doi.org/10.1016/j.bbi.2020.05.019)] [Medline: [32389702](https://pubmed.ncbi.nlm.nih.gov/32389702/)]
54. Viswanath K, Nagler RH, Bigman-Galimore CA, McCauley MP, Jung M, Ramanadhan S. The communications revolution and health inequalities in the 21st Century: implications for cancer control. *Cancer Epidemiol Biomarkers Prev* 2012 Oct 07;21(10):1701-1708. [doi: [10.1158/1055-9965.epi-12-0852](https://doi.org/10.1158/1055-9965.epi-12-0852)]
55. Molina Y, Zimmermann K, Carnahan LR, Paulsey E, Bigman CA, Khare MM, et al. Rural women's perceptions about cancer disparities and contributing factors: a call to communication. *J Cancer Educ* 2018 Aug 27;33(4):749-756 [FREE Full text] [doi: [10.1007/s13187-017-1196-5](https://doi.org/10.1007/s13187-017-1196-5)] [Medline: [28243956](https://pubmed.ncbi.nlm.nih.gov/28243956/)]
56. Jensen JD, King AJ, Carcioppolo N, Krakow M, Samadder NJ, Morgan S. Comparing tailored and narrative worksite interventions at increasing colonoscopy adherence in adults 50-75: a randomized controlled trial. *Soc Sci Med* 2014 Mar;104:31-40. [doi: [10.1016/j.socscimed.2013.12.003](https://doi.org/10.1016/j.socscimed.2013.12.003)] [Medline: [24581059](https://pubmed.ncbi.nlm.nih.gov/24581059/)]
57. Han PKJ, Moser RP, Klein WMP. Perceived ambiguity about cancer prevention recommendations: associations with cancer-related perceptions and behaviours in a US population survey. *Health Expect* 2007 Dec;10(4):321-336 [FREE Full text] [doi: [10.1111/j.1369-7625.2007.00456.x](https://doi.org/10.1111/j.1369-7625.2007.00456.x)] [Medline: [17986069](https://pubmed.ncbi.nlm.nih.gov/17986069/)]
58. Niederdeppe J, Levy AG. Fatalistic beliefs about cancer prevention and three prevention behaviors. *Cancer Epidemiology Biomarkers & Prevention* 2007 May 01;16(5):998-1003. [doi: [10.1158/1055-9965.epi-06-0608](https://doi.org/10.1158/1055-9965.epi-06-0608)]
59. Farooq A, Laato S, Islam AKMN. Impact of online information on self-isolation intention during the COVID-19 pandemic: cross-sectional study. *J Med Internet Res* 2020 May 06;22(5):e19128 [FREE Full text] [doi: [10.2196/19128](https://doi.org/10.2196/19128)] [Medline: [32330115](https://pubmed.ncbi.nlm.nih.gov/32330115/)]
60. Siebenhaar KU, Köther AK, Alpers GW. Dealing with the COVID-19 infodemic: distress by information, information avoidance, and compliance with preventive measures. *Front Psychol* 2020 Nov 5;11:567905 [FREE Full text] [doi: [10.3389/fpsyg.2020.567905](https://doi.org/10.3389/fpsyg.2020.567905)] [Medline: [33224060](https://pubmed.ncbi.nlm.nih.gov/33224060/)]
61. Pickles K, Cvejic E, Nickel B, Copp T, Bonner C, Leask J, et al. COVID-19 misinformation trends in Australia: prospective longitudinal national survey. *J Med Internet Res* 2021 Jan 07;23(1):e23805 [FREE Full text] [doi: [10.2196/23805](https://doi.org/10.2196/23805)] [Medline: [33302250](https://pubmed.ncbi.nlm.nih.gov/33302250/)]
62. Voigt T, Holtz V, Niemiec E, Howard H, Middleton A, Prainsack B. Willingness to donate genomic and other medical data: results from Germany. *Eur J Hum Genet* 2020 Aug;28(8):1000-1009 [FREE Full text] [doi: [10.1038/s41431-020-0611-2](https://doi.org/10.1038/s41431-020-0611-2)] [Medline: [32238912](https://pubmed.ncbi.nlm.nih.gov/32238912/)]
63. Baker R, Blumberg SJ, Brick JM, Couper MP, Courtright M, Dennis JM, et al. Research Synthesis: AAPOR Report on Online Panels. *Public Opinion Quarterly* 2010 Oct 20;74(4):711-781. [doi: [10.1093/poq/nfq048](https://doi.org/10.1093/poq/nfq048)]
64. Blom AG, Gathmann C, Krieger U. Setting up an online panel representative of the general population. *Field Methods* 2015 Mar 26;27(4):391-408. [doi: [10.1177/1525822x15574494](https://doi.org/10.1177/1525822x15574494)]
65. Stanley M, Roycroft J, Amaya A, Dever JA, Srivastava A. The effectiveness of incentives on completion rates, data quality, and nonresponse bias in a probability-based internet panel survey. *Field Methods* 2020 Feb 18;32(2):159-179. [doi: [10.1177/1525822x20901802](https://doi.org/10.1177/1525822x20901802)]
66. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *J Med Internet Res* 2004 Sep 29;6(3):e34 [FREE Full text] [doi: [10.2196/jmir.6.3.e34](https://doi.org/10.2196/jmir.6.3.e34)] [Medline: [15471760](https://pubmed.ncbi.nlm.nih.gov/15471760/)]
67. Lee S, Yuen S, Or N, Cheng EW, Yue R. Pandemic vulnerability, policy feedback and support for immigration: Evidence from Asia. *Br J Soc Psychol* 2022 Mar 04;1-20. [doi: [10.1111/bjso.12529](https://doi.org/10.1111/bjso.12529)] [Medline: [35244226](https://pubmed.ncbi.nlm.nih.gov/35244226/)]
68. Weathers F, Litz B, Keane T, Palmieri P, Marx B, Schnurr P. PTSD Checklist for DSM-5 (PCL-5). U.S. Department of Veterans Affairs. 2013 Aug 19. URL: <https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp> [accessed 2022-02-25]
69. Sarkhel S, Bakhla A, Praharaj S, Ghosal M. Information overload regarding COVID-19: Adaptation and validation of the cancer information overload scale. *Indian J Psychiatry* 2020;62(5):481. [doi: [10.4103/psychiatry.indianjpsychiatry_974_20](https://doi.org/10.4103/psychiatry.indianjpsychiatry_974_20)]
70. Schmidt A, Finan C. Linear regression and the normality assumption. *J Clin Epidemiol* 2018 Jun;98:146-151 [FREE Full text] [doi: [10.1016/j.jclinepi.2017.12.006](https://doi.org/10.1016/j.jclinepi.2017.12.006)] [Medline: [29258908](https://pubmed.ncbi.nlm.nih.gov/29258908/)]
71. Hirschler B. Information overload. Brunswick Group. 2021 Mar 29. URL: <https://www.brunswickgroup.com/scientists-publishing-risks-i18546/> [accessed 2022-02-25]
72. Bonoil M, McIsaac M, Xu L, Wuliji T, Diallo K, Campbell J. Gender equity in the health workforce: analysis of 104 countries. World Health Organization. 2019. URL: <https://apps.who.int/iris/bitstream/handle/10665/311314/WHO-HIS-HWF-Gender-WP1-2019.1-eng.pdf> [accessed 2022-02-25]

73. Galasso V, Pons V, Profeta P, Becher M, Brouard S, Foucault M. Gender differences in COVID-19 attitudes and behavior: Panel evidence from eight countries. *Proc Natl Acad Sci U S A* 2020 Nov 03;117(44):27285-27291 [FREE Full text] [doi: [10.1073/pnas.2012520117](https://doi.org/10.1073/pnas.2012520117)] [Medline: [33060298](https://pubmed.ncbi.nlm.nih.gov/33060298/)]
74. Zhao X, Fan J, Basnyat I, Hu B. Online health information seeking using "#COVID-19 Patient Seeking Help" on Weibo in Wuhan, China: descriptive study. *J Med Internet Res* 2020 Oct 15;22(10):e22910 [FREE Full text] [doi: [10.2196/22910](https://doi.org/10.2196/22910)] [Medline: [33001838](https://pubmed.ncbi.nlm.nih.gov/33001838/)]
75. Lee K, Hoti K, Hughes JD, Emmerton LM. Consumer Use of "Dr Google": a survey on health information-seeking behaviors and navigational needs. *J Med Internet Res* 2015 Dec 29;17(12):e288 [FREE Full text] [doi: [10.2196/jmir.4345](https://doi.org/10.2196/jmir.4345)] [Medline: [26715363](https://pubmed.ncbi.nlm.nih.gov/26715363/)]
76. Zickuhr K. Generations 2010. Pew Internet Research. 2010 Dec 16. URL: https://www.pewresearch.org/internet/wp-content/uploads/sites/9/2010/12/PI_2010.12.16_Generations-and-Tech10_FINAL.pdf [accessed 2022-02-25]
77. Fox S, Rainie L. E-patients and the online health care revolution. *Physician Exec* 2002;28(6):14-17. [Medline: [12448136](https://pubmed.ncbi.nlm.nih.gov/12448136/)]
78. Chae J, Lee C, Kim K. Prevalence, predictors, and psychosocial mechanism of cancer information avoidance: findings from a national survey of U.S. adults. *Health Commun* 2020 Mar 04;35(3):322-330. [doi: [10.1080/10410236.2018.1563028](https://doi.org/10.1080/10410236.2018.1563028)] [Medline: [30606065](https://pubmed.ncbi.nlm.nih.gov/30606065/)]
79. Rubin GJ, Amlôt R, Page L, Wessely S. Public perceptions, anxiety, and behaviour change in relation to the swine flu outbreak: cross sectional telephone survey. *BMJ* 2009 Jul 02;339(jul02 3):b2651-b2651 [FREE Full text] [doi: [10.1136/bmj.b2651](https://doi.org/10.1136/bmj.b2651)] [Medline: [19574308](https://pubmed.ncbi.nlm.nih.gov/19574308/)]
80. Soroya SH, Farooq A, Mahmood K, Isoaho J, Zara S. From information seeking to information avoidance: Understanding the health information behavior during a global health crisis. *Inf Process Manag* 2021 Mar;58(2):102440 [FREE Full text] [doi: [10.1016/j.ipm.2020.102440](https://doi.org/10.1016/j.ipm.2020.102440)] [Medline: [33281273](https://pubmed.ncbi.nlm.nih.gov/33281273/)]
81. Trumbo CW. Information processing and risk perception: An adaptation of the heuristic-systematic model. *Journal of Communication* 2002;52(2):367-382. [doi: [10.1111/j.1460-2466.2002.tb02550.x](https://doi.org/10.1111/j.1460-2466.2002.tb02550.x)]
82. Talev M. Axios-Ipsos coronavirus index: Rich sheltered, poor shafted amid virus. *Axios*. 2020 Apr 01. URL: <https://www.axios.com/axios-ipsos-coronavirus-index-rich-sheltered-poor-shafted-9e592100-b8e6-4dcd-ae0-a6516892874b.html> [accessed 2022-02-25]
83. Chen X, Gao H, Zou Y, Lin F. Changes in psychological wellbeing, attitude and information-seeking behaviour among people at the epicentre of the COVID-19 pandemic: a panel survey of residents in Hubei province, China. *Epidemiol Infect* 2020 Sep 02;148:e201 [FREE Full text] [doi: [10.1017/S0950268820002009](https://doi.org/10.1017/S0950268820002009)] [Medline: [32873358](https://pubmed.ncbi.nlm.nih.gov/32873358/)]
84. Klerings I, Weinhandl AS, Thaler KJ. Information overload in healthcare: too much of a good thing? *Z Evid Fortbild Qual Gesundheitswes* 2015;109(4-5):285-290. [doi: [10.1016/j.zefq.2015.06.005](https://doi.org/10.1016/j.zefq.2015.06.005)] [Medline: [26354128](https://pubmed.ncbi.nlm.nih.gov/26354128/)]
85. Lee T, Lee B, Lee-Geiller S. The effects of information literacy on trust in government websites: Evidence from an online experiment. *International Journal of Information Management* 2020 Jun;52:102098. [doi: [10.1016/j.ijinfomgt.2020.102098](https://doi.org/10.1016/j.ijinfomgt.2020.102098)]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
DSM-5: Diagnostic and Statistical Manual of Mental Disorders 5
MERS: Middle East Respiratory Syndrome
OLS: ordinary least squares
PCL-5: Posttraumatic Stress Disorder Checklist for DSM-5
SARS: Severe Acute Respiratory Syndrome
SES: socioeconomic status
WHO: World Health Organization

Edited by C Basch; submitted 08.06.21; peer-reviewed by K Kishore, M Salimi, T Purnat; comments to author 05.07.21; revised version received 17.08.21; accepted 13.10.21; published 22.03.22.

Please cite as:

Chen X, Lin F, Cheng EW

Stratified Impacts of the Infodemic During the COVID-19 Pandemic: Cross-sectional Survey in 6 Asian Jurisdictions

J Med Internet Res 2022;24(3):e31088

URL: <https://www.jmir.org/2022/3/e31088>

doi: [10.2196/31088](https://doi.org/10.2196/31088)

PMID: [35103601](https://pubmed.ncbi.nlm.nih.gov/35103601/)

©Xi Chen, Fen Lin, Edmund W Cheng. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 22.03.2022. 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.

Original Paper

The Evaluation of a Social Media Campaign to Increase COVID-19 Testing in Migrant Groups: Cluster Randomized Trial

Ingeborg Hess Elgersma¹, MA; Atle Fretheim^{1,2}, MD, PhD; Thor Indseth³, MA; Anita Thorolvsen Munch⁴, MA; Live Bøe Johannessen⁴, MA; Christine Engh Hansen⁵, BA

¹Centre for Epidemic Interventions Research, Norwegian Institute of Public Health, Oslo, Norway

²Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway

³Cluster for Health Services Research, Norwegian Institute of Public Health, Oslo, Norway

⁴Division of Prevention and Public Health, Norwegian Directorate of Health, Oslo, Norway

⁵Mindshare Norway, Oslo, Norway

Corresponding Author:

Ingeborg Hess Elgersma, MA

Centre for Epidemic Interventions Research

Norwegian Institute of Public Health

Postboks 222, Skøyen

Oslo, 0213

Norway

Phone: 47 91584782

Email: ingeborghess.elgersma@fhi.no

Abstract

Background: A low test positivity rate is key to keeping the COVID-19 pandemic under control. Throughout the pandemic, several migrant groups in Norway have seen higher rates of confirmed COVID-19 and related hospitalizations, while test positivity has remained high in the same groups. The Norwegian government has used several platforms for communication, and targeted social media advertisements have in particular been an important part of the communication strategy to reach these groups.

Objective: In this study, we aimed to investigate whether such a targeted Facebook campaign increased the rate of COVID-19 tests performed in certain migrant groups.

Methods: We randomly assigned 386 Norwegian municipalities and city districts to intervention or control groups. Individuals born in Eritrea, Iraq, Pakistan, Poland, Russia, Somalia, Syria, and Turkey residing in intervention areas were targeted with a social media campaign aiming at increasing the COVID-19 test rate. The campaign message was in a simple language and conveyed in the users' main language or in English.

Results: During the 2-week follow-up period, the predicted probability of having a COVID-19 test taken was 4.82% (95% CI 4.47%-5.18%) in the control group, and 5.58% (95% CI 5.20%-5.99%) in the intervention group ($P=.004$).

Conclusions: Our targeted social media intervention led to a modest increase in test rates among certain migrant groups in Norway.

Trial Registration: ClinicalTrials.gov NCT04866589; <https://clinicaltrials.gov/ct2/show/NCT04866589>

(*J Med Internet Res* 2022;24(3):e34544) doi:[10.2196/34544](https://doi.org/10.2196/34544)

KEYWORDS

COVID-19; SARS-CoV-2; social media; campaign; cluster randomized trial; nonpharmaceutical interventions; migrant; intervention; testing; strategy; public health; Facebook; communication

Introduction

Several migrant groups have been disproportionately affected by COVID-19 in Norway [1,2]. These groups are also considered difficult to reach through traditional communication

platforms [3,4]. Adapting, translating, and targeting communication to ensure equal access to information regardless of health literacy and language skills, is an important part of crisis communication. Consequently, Norwegian health authorities have used a wide range of platforms to communicate with these groups during the pandemic: public COVID-19

websites, traditional media outlets such as television, radio, boards and printed advertisements, as well as content marketing, contextual digital ads, targeted cell phone messaging, influencers, and a wide array of social media such as Twitter, Instagram, Snapchat, TikTok, YouTube, and Facebook. Certain migrant groups have seen particularly high COVID-19 prevalence and low test rates [5], and social media campaigns can be crucial to reach these groups. Although such campaigns have been monitored in house in terms of outreach and reactions on the social media platform, their effects on behavioral change have not been assessed.

Testing is key to detecting cases and limiting outbreaks. In this study, we aimed at assessing whether a targeted social media advertising campaign encouraging selected migrant groups to get tested affected testing behavior (ClinicalTrials.gov NCT04866589; protocol is provided in [Multimedia Appendices 1 and 2](#)). The intervention described in this paper was an extension of the “Get tested” campaign, which had been launched after the Norwegian Institute of Public Health (NIPH) published findings that indicated higher rates of undetected cases among certain migrant groups compared to the general population; that is, lower test rates, higher test positivity rates, and higher rates of hospitalization [1]. The “Get tested” campaign was a part of a larger effort directed at the migrant population, which included drop-in testing stations [6] and dialogue meetings with migrant groups.

The campaign was launched by the end of May 2021, when the COVID-19 incidence rate in Norway was decreasing, but the test positivity rate—that is, the share of all tests that are positive—remained high in many migrant groups [5]. Furthermore, there were indications of confusion concerning when to test and whether testing was free. In order to assess the campaign’s effect on the rates of testing, we ran a cluster randomized trial with users of Facebook. According to numbers from the Ipsos SoMe-tracker, 82% of the Norwegian population, aged ≥ 18 years, have a profile, and 69% of them are daily users on this social media platform [7]. The campaign targeted users residing in Norway, who were born in Eritrea, Iraq, Pakistan, Poland, Russia, Somalia, Syria, and Turkey. The groups were chosen because of their previous high test positivity levels. This study aimed to establish whether the targeted Facebook campaign increased the rate of testing in the selected migrant groups.

Methods

Trial Design

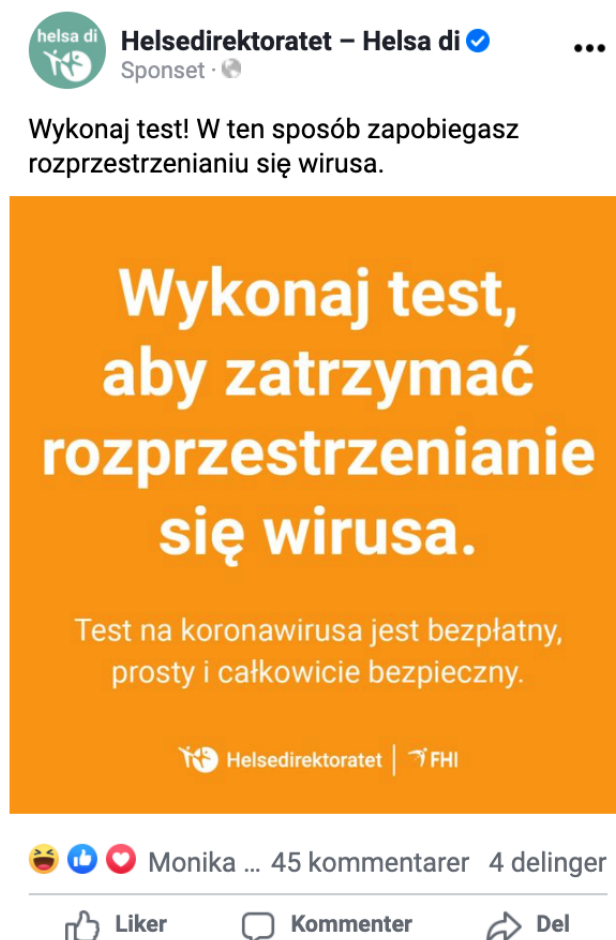
The trial was designed as a parallel group, 2-arm, superiority cluster randomized trial. The clusters—that is, municipalities and city districts—were allocated either to the intervention

group or to the control group, where the intervention group received the campaign message and the control group did not receive the campaign message. All municipalities or city districts with at least one person in the target group were included in the study. Randomization was carried out at the cluster level as it was not technically possible to target the campaign to individual persons. The clusters were split evenly between treatments through block randomization with block size two. Before block randomization took place, the list of municipalities and city districts were sorted in accordance with the total number of migrants residing in each area. Partial concealment was ensured as team member 1 created and sorted the list, while team number 2 created a list of pairwise random orders of pairs of A and B, using randomization software [8]. Team number 3 combined these two lists.

Intervention

Norwegian health authorities have, since the onset of the pandemic, carried out campaigns on social media directed toward the migrant population, using language-based and geographic targeting segments. In December 2020 a “Get tested” campaign was launched, aiming to increase the test rate among migrant populations. Prior to our trial, there were indications of an increased disinterest toward the message, as measured by reach and reactions, and the campaign was paused. To prevent message fatigue [9,10], campaign messages were continuously reviewed, performance measured and customized, and A/B testing conducted—this involved testing different colors, visuals, or creatives to determine which visual received more attention and the time spent on the message. As the “Get tested” campaign had been running for a while, there was by the end of May a small concern that the public had become disengaged or blind to the message; hence, it was decided to roll out new versions. The new versions used short, direct messaging on an A/B-tested colored background. The message remained the same.

We decided that the distribution of the new campaign would be randomized by municipality or city district in order to facilitate an effect evaluation. The content of the relaunched campaign was a sponsored Facebook post from the Norwegian Directorate of Health, distributed in Tigrinya, Polish, Urdu, Somali, Russian, Kurmanji, English, Arabic, Turkish, and Sorani, and focused on the importance of testing for COVID-19 and that testing is free of charge and easy to take. The English message read the following: “COVID-19 tests are free, simple, and completely safe. Get yourself tested to stop the spread of the infection.” Facebook users were free to comment under the post and to share it. [Figure 1](#) shows a screenshot of the post in Polish. Facebook was chosen as a platform in particular because of the high number of Facebook users in Norway in general, and because this platform allowed us to target the position of users with high accuracy.

Figure 1. Screenshot of the social media post in Polish.

The campaign was disseminated using targeted Facebook Ads and was launched separately for each municipality or city district and each language. Target groups on Facebook were created by specifying the geographic position and language setting. It was not possible to launch a Facebook campaign when fewer than 100 users matched the targeting criteria. In cases where not enough users matched the language setting, target groups were formed on the basis of behavioral attributes such as “living abroad,” and interest segments specific to the users’ native country. Examples of the latter are as follows: “Syria national football team,” “BBC Urdu,” and “Turkish Kurdistan.” At the time of the trial, Somali and Tigrinya were not available as languages on Facebook. Users who spoke these languages were solely targeted on the basis of behavioral attributes. Lastly, the English campaign was disseminated on the basis of the attribute “living abroad” and the English user language setting.

The campaign was live for a 7-14-day period. The variation in timing was due to the time needed to roll out all the campaigns. Of the 117,436 participants in the intervention group, 52,565 lived in areas where it was possible to target the campaign in accordance with Facebook’s language settings, 58,336 lived in areas where it was only possible to target the campaign on the basis of behavioral attributes or interests, and 4700 lived in areas where it was not possible to launch the campaign, neither based on the participants’ background or language. We cannot know for certain to what extent the intervention group, living

in areas where the campaign was rolled out in their language, was actually exposed to the intervention.

Participants

The sample included 233,903 persons with a Norwegian national identity number, who are registered as a resident of a Norwegian municipality or city district (in Oslo, Bergen, and Trondheim), aged 18 years or older at the time of assignment, and registered are as having been born in Eritrea, Iraq, Pakistan, Poland, Russia, Somalia, Syria, or Turkey.

Data Source and Outcomes

We used registry data from the emergency preparedness register for COVID-19 (Beredit C19), which was established by the NIPH in April 2020 [11]. The register contains data on all individuals in the Norwegian Population Register. Individuals were linked across data sources and over time, using an encrypted version of the unique personal identification number provided for every resident of Norway at birth or upon first immigration. In this study, we used individual data from the Norwegian Surveillance System for Communicable Diseases (MSIS) and the National Population Register. MSIS contains data on all polymerase chain reaction (PCR) and rapid tests taken in Norway at official test stations.

Data were also obtained from Facebook on whether it was possible to roll out the campaign, and on the reach, exposure, and frequency of the campaigns that were run in each cluster. These data were anonymized and used only in descriptive

analyses, as data from Facebook cannot be linked directly to data on testing.

Data were aggregated into two time periods: baseline data included data collected 14 days to 1 day before the campaign commenced. Outcome data were measured from day 1 to day 14 and from day 1 to day 21 after the campaign started. Day zero was excluded from the analysis. The exclusion of day zero was prespecified in the protocol, and the rationale was that it was unlikely that the effects of the campaign could be observed on the first day.

The outcome was a dichotomous variable, measured at the individual level, indicating whether a person had taken a PCR or antigen test for SARS-CoV-2.

Time invariant covariates included country of birth, sex, and age. These variables were obtained from the National Population Register.

Statistical Analysis

The coefficient of interest was β_1 in regression analysis:

$$y_{j|i} = \alpha_{j|i} + \beta_1 \text{Treated}_i + \beta_2 y_{i0} + \mathbf{X}_i \beta_3 + \varepsilon_{j|i},$$

where $y_{j|i}$ indicates whether the individual i in municipality or city district j has taken a COVID-19 PCR or rapid antigen test during the 2 weeks the campaign lasted. y_{i0} is the baseline value indicating whether the individual i has taken a COVID-19 test in the 2 weeks prior to campaign commencement. We adjust for the baseline value as is customary in randomized controlled trials (RCTs) [12,13]. \mathbf{X}_i is a vector of control variables.

As the outcome is binary, measured at the individual level, and the treatment was assigned at the cluster level, j , multilevel logistic regression analysis was performed to account for intraclass correlation.

The main specification does not take into account that it was not possible to administer the intervention in every municipality or city district assigned to the intervention group. As such, the results represent an “intention to treat” estimate.

Sample Size Estimation

The sample size relied on the number of persons aged 18 years and older, born in Eritrea, Pakistan, Poland, Russia, Somalia, Syria, or Turkey, and residing in Norway. The number of languages and countries was determined by the availability of campaign material in the different languages, the distribution of these groups across Norway, and the size of the groups. The language groups were also selected on the grounds of a higher test positivity rate as opposed to the general population. Thus, our sample consisted of all eligible individuals, not a selected sample. Consequently, we performed no prior sample size estimation.

Ethics and Privacy Issues

This study does not qualify as health research in the legal sense and hence does not need formal ethics approval. Participant

consent was waived because this study was an evaluation of an intervention, using anonymous registry data already collected.

Personal data protection was ensured through the rigorous set up of the Beredt C19 where data on testing and country of birth were linked to the encrypted version of the unique personal identification number. Data from Facebook were collected and aggregated by Mindshare, and could not be linked directly to the registry data.

Results

The campaign was live from June 1 to June 14, 2021. In 119 out of 189 municipalities or city districts, the campaign was rolled out on June 1, while in the remaining 70 municipalities, the campaign was live from June 6, 2021. In total, the campaigns reached more than 351,000 individual users. The users were likely exposed to the same campaign several times but in a different language, which was expected, as there were multiple languages per target country. The number of individual users far exceeds the 233,903 persons included in our target group, which likely entails that it also reached other users; for example, nonmigrant users, nonmigrants born to migrant parents, or migrants not born in 1 of the 8 target countries.

Summary statistics on the sample that was randomized are shown in Table 1 (Figure S1 in Multimedia Appendix 3 shows their localization on the map). The different migrants were quite evenly distributed between the intervention and control groups. In total, 42.7% of individuals in the intervention group and 42.2% of those in the control group were female. The average age was 41 years in the intervention group and 42 years in the control group. The average cluster sizes were similar in the intervention and control groups.

A striking difference between the groups is the preintervention COVID-19 fortnightly incidence rate, indicating that the randomization process did not yield a completely balanced sample. The number of fortnightly cases was much higher in the intervention group (123.6 cases per 100,000 population) than in the control group (79.6 cases per 100,000 population). This difference is likely to have influenced postoutcome testing, as the test-isolate-trace-quarantine strategy implies that a high number of persons will be tested in conjunction with COVID-19 outbreaks. In Figure S2 in Multimedia Appendix 3, we show that the higher number of cases is driven by a few observations, notably outbreaks in Hammerfest and Larvik municipalities. We therefore controlled for whether the individual had taken a COVID-19 test during the baseline period, and conducted a separate sensitivity analysis where we excluded the two pairs of clusters with the biggest preintervention differences in COVID-19 incidence.

Figure 2 shows the proportion tested daily in the intervention and control groups. The proportion of individuals tested daily in the intervention group was higher than that in the control group on most days both before and after the intervention was carried out.

Table 1. Baseline demographic characteristics of participants and clusters by experimental arm (N=233,903).

Variable	Intervention group (n=117,436)	Control group (n=116,467)
Individual-level characteristics		
Gender, n (%)		
Female	50,194 (42.7)	49,176 (42.2)
Male	67,242 (57.3)	67,291 (57.8)
Country of birth, n (%)		
Eritrea	9562 (8.1)	9662 (8.3)
Iraq	11,395 (9.7)	10,256 (8.8)
Pakistan	12,153 (10.3)	8387 (7.2)
Poland	45,319 (38.6)	49,247 (42.3)
Russia	8107 (6.9)	9204 (7.9)
Somalia	12,575 (10.7)	12,192 (10.5)
Syria	11,425 (9.7)	12,057 (10.4)
Turkey	6900 (5.9)	5462 (4.7)
Age, mean (SD)	41 (12)	40 (12)
Cluster-level characteristics		
Clusters, n	191	191
Cluster size, mean (SD; range)	609.8 (1154.5; 5-6931)	614.8 (1186.8; 2-7804)
Population, n (mean)	2,858,940 (14,968.3)	2,649,940 (13,874.0)
COVID-19 baseline fortnightly incidence (per 100,000)	79.6	123.6

Figure 2. The daily proportion of the participants tested in the control and in the intervention groups. Error bars represent 95% CIs.

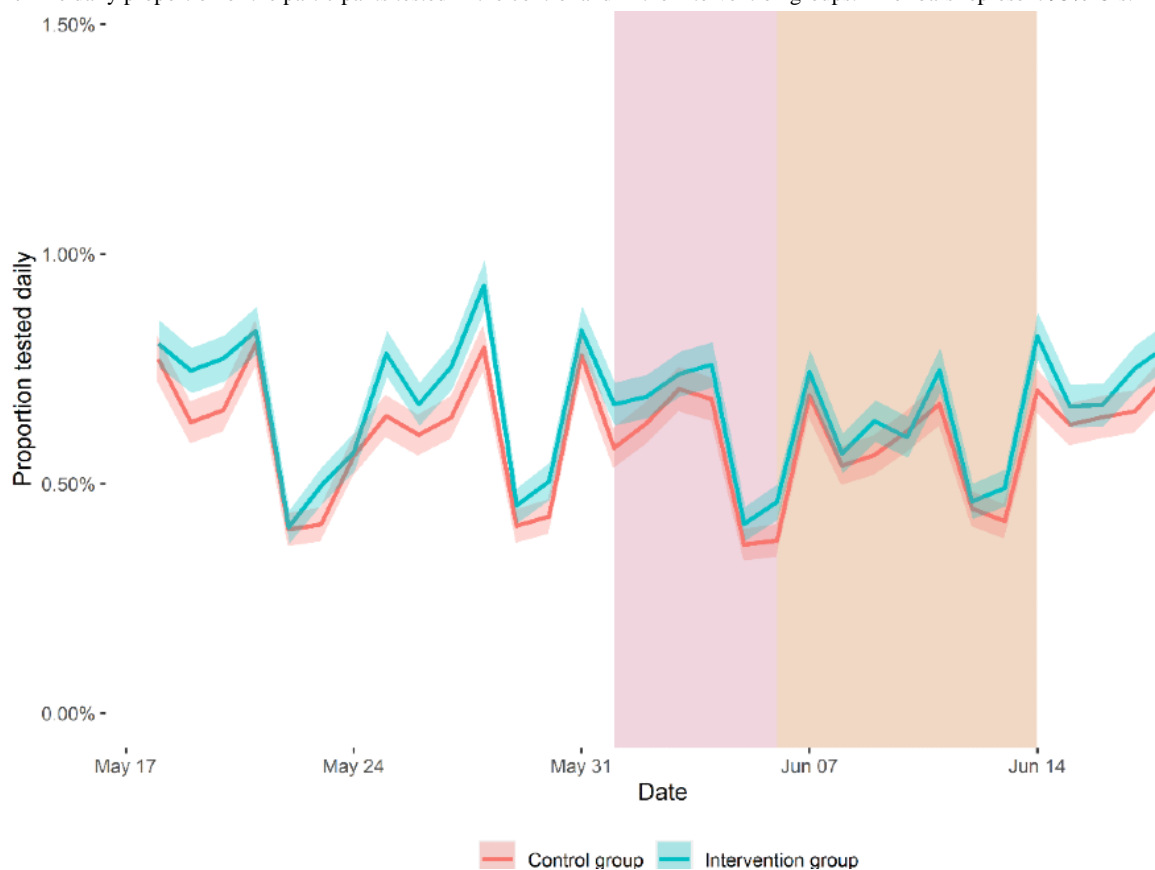


Table 2 and Table S2 in the [Multimedia Appendix 4](#) display the results of regression analysis. The odds ratio (OR) for the intervention group was 1.17 (95% CI 1.05-1.30; $P=.004$); that is, the campaign had a significant positive effect on the likelihood of undergoing a COVID-19 test. Whether a person had taken a COVID-19 test during the baseline period was a strong predictor of whether that person would take a test during the follow-up period (OR 5.34, 95% CI 5.12-5.56; $P=.001$). Controlling for age and sex did not affect the effect estimate.

Figure 3 shows the predicted probability of undergoing a COVID-19 test in the 14-day follow-up period—this was 4.82% (95% CI 4.47%-5.18%) in the control group (areas without exposure to the campaign) and 5.58% (95% CI 5.20%-5.99%) in the intervention group (areas with exposure to the campaign). The 0.76 percentage point difference between the control and intervention groups converts to a 15.7% relative increase in test rates attributed to the campaign.

Table S1 in the [Multimedia Appendix 3](#) shows the effects of the campaign when the follow-up period was 1-21 days after the campaign commenced rather than 1-14 days considered for our main analysis. The effect on testing remained significant (OR 1.13, 95% CI 1.02-1.25; $P=.02$). The results are also robust to the exclusion of the two pairs with the largest preintervention (baseline) differences in COVID-19 incidence. When we examined whether the effects differed in accordance with which targeting criteria was used to segment target groups, the results show that targeting based on the users' language settings yielded a larger effect (OR 1.29, 95% CI 1.16-1.44; $P<.001$) than that based on behavioral attributes or interest (OR 1.11, 95% CI 1.00-1.24; $P=.045$). Not surprisingly, there was no significant difference in testing rates between individuals in the control group and those in areas where it was not possible to launch the campaign (OR 0.99, 95% CI 0.83-1.18; $P=.89$).

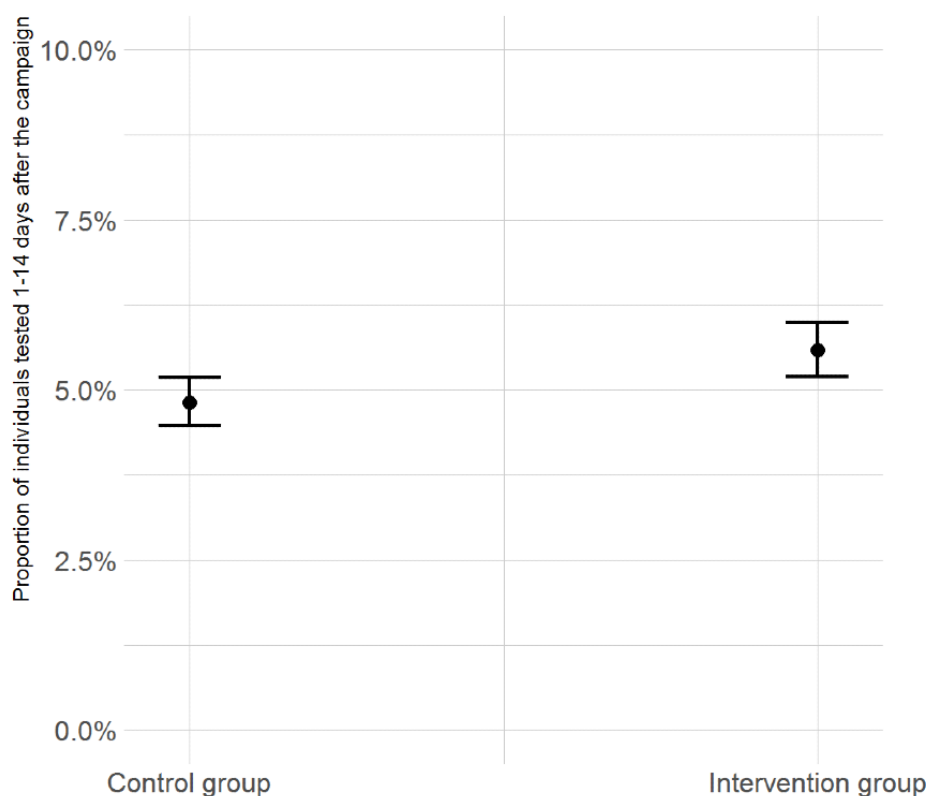
Table 2. Results of regression analysis.

Predictors	Tested (1-14 days) ^a		Tested (1-14 days) ^b	
	Odds ratio (95% CI)	<i>P</i> value	Odds ratio (95% CI)	<i>P</i> value
Intercept	0.04 (0.04-0.05)	<.001	0.04 (0.03-0.04)	<.001
Intervention group	1.17 (1.05-1.30)	.004	1.17 (1.05-1.30)	.004
Tested precampaign	5.34 (5.12-5.56)	<.001	5.24 (5.03-5.46)	<.001
Female	N/A ^c	N/A	1.25 (1.21-1.30)	<.001
Age (years)				
25-39	N/A	N/A	1.13 (1.06-1.21)	<.001
40-44	N/A	N/A	1.12 (1.04-1.21)	.003
45-49	N/A	N/A	1.18 (1.09-1.27)	<.001
50-59	N/A	N/A	1.11 (1.03-1.19)	.007
60-70	N/A	N/A	1.02 (0.93-1.11)	.68
>70	N/A	N/A	0.65(0.56-0.76)	<.001

^aModel adjusted only for baseline value.

^bModel adjusted for baseline value, gender, and age.

^cN/A: not applicable.

Figure 3. Predicted probabilities of conducting a COVID-19 test in the 14-day follow-up period. Error bars represent 95% CIs.

Discussion

Principal Findings

We conducted this trial to assess the effect of a targeted social media campaign on the rates of testing among Norwegian residents born in Eritrea, Iraq, Pakistan, Poland, Russia, Somalia, Syria, or Turkey. We found a 15.7% relative increase in the proportion taking a COVID-19 test among those who were exposed to the campaign, or an absolute increase of 0.76 percentage points. A distinctive feature of this trial, that strengthens the significance of our findings, was that we were able to assess the effects on actual behavior and not only on intentions or beliefs [14].

We found that targeting in accordance with the users' language setting yielded a larger effect than that based on behavioral attributes or interest. The content and message remained the same for both targeting approaches. A probable explanation is that the targeting based on users' language settings reached a higher share of the participants than that based on behavioral attributes or interests.

Comparison With Prior Work

We are aware of some trials with a similar objective, assessing the effects of social media campaigns on COVID-19–related preventive behavior, COVID-19 knowledge, and intent toward COVID-19 prevention.

We identified 3 RCTs that test the effect of social media campaigns on COVID-19 prevention knowledge. Alsan et al [15] examined how messages read by physicians who varied in age, gender, race, and ethnicity influenced knowledge, beliefs, and practices related to COVID-19. Physician-delivered

messages increased knowledge of COVID-19 symptoms and prevention methods for Black and Latinx respondents. The results illustrate that tailoring the message to the targeted groups may increase the effectiveness of the interventions. Similarly, Vandormael et al [16] have conducted a web-based RCT to investigate the effect of a short video on improving COVID-19 prevention knowledge and behavioral intent toward COVID-19 prevention. They concluded the following: “Short, wordless, animated videos, distributed by health authorities via social media, may be an effective pathway for rapid global health communication during health crises” [16]. Lastly, the results of Alatas et al [17] suggest that celebrity endorsement in a social media campaign in Indonesia influenced beliefs about vaccination and knowledge of immunization.

Only Breza et al [18] investigated how social media advertisements affect actual COVID-19 preventive behavior. They showed that short messages recorded by health professionals before the winter holidays in the United States and sent as advertisements to social media users led to reductions in travel (–0.993 percentage points in high-intensity counties, $P=.002$) and a decrease in COVID-19 infection at the zip code level in the 2-week period starting 5 days post holiday (3.5%, $P=.01$) [18].

Our results are aligned with those of the aforementioned studies. We demonstrated that targeted advertisements on social media with short, simplified messages can be a valuable tool in the authorities' toolbox in times of a pandemic. Targeted advertisements can reach audiences that are not easily reached by traditional modes of communication. Furthermore, conveying the message in the mother tongue of the audience can increase trust in the message, although we did not test this specifically.

The effect size was small, which was not unexpected, considering that the COVID-19 prevalence was quite low in the weeks before campaign onset, and that the intervention was a part of a longer ongoing campaign. Norway saw a large increase in test rates among migrants toward the end of 2020 and in early 2021, after the “Get tested” campaign was launched and amid media attention on the high incidence of COVID-19 among migrant groups [19]. It is therefore likely that toward the end of 2021, the target groups were experiencing a certain advertisement fatigue, and that the main message of the campaign was well known both in the control group and the intervention group.

Limitations

There are some limitations associated with the findings of this study. First, the design of the trial limited the scope and content of the campaign, and the likely external validity. Because we needed to roll out separate campaigns in all areas, it was too arduous to also put out videos or other types of campaign material. In a real-life setting, it would not be necessary to limit users to a very specific geographic unit, the costs of administering the campaign would be lower, and there would be room for other modes of communication.

Another limitation was that it was not possible to identify the exact effect of actually seeing the posts on testing behavior, as data on reach could not be directly linked with registry data on testing. We know that 82% of the population aged 18 years or

older has a Facebook profile [7], but this is only one of many social media platforms. Nevertheless, the total number of users speaking different languages, who were reached, suggests that a large share of the target group was exposed to the campaign.

A related limitation is that the campaign may have had spillover effects on the control group; for example, persons in the intervention group may have shared the post to their Facebook friends and followers. Such organic reach could not be measured, nor was it avoidable. Similarly, the message could have spread from the intervention group to the control group by word of mouth. Spillover effects may have led to an underestimation of the true effect.

Future research could look into effects of communication campaigns on other types of COVID-19 preventive behavior. We also suggest that future campaigns consider using the emergency broadcast capability inherent in cell phone services to deliver targeted advertisements.

Conclusions

Seeing a native-language post on Facebook, in a clear language easily understood manner regardless of health literacy, explaining that testing is simple and can be taken at no cost rendered our target group more likely to take a COVID-19 test. This study demonstrated that targeted social media advertisements sponsored by health authorities can influence individual behavior in an infection control–friendly direction during a pandemic.

Acknowledgments

We would like to thank Gutorm Høgåsen, Ragnhild Tønnessen, and Anja Elsrud Schou Lindman for their invaluable efforts in maintaining the Beredt C19 Registry.

Conflicts of Interest

CEH is an employee at Mindshare, a marketing company that sells social media-related services.

Multimedia Appendix 1

Protocol: Evaluation of social media campaign to increase COVID-19 testing in migrant groups: A cluster randomised trial.
[DOCX File, 40 KB - [jmir_v24i3e34544_app1.docx](#)]

Multimedia Appendix 2

CONSORT extension for Cluster Trials.
[DOCX File, 33 KB - [jmir_v24i3e34544_app2.docx](#)]

Multimedia Appendix 3

Additional results.
[DOCX File, 457 KB - [jmir_v24i3e34544_app3.docx](#)]

Multimedia Appendix 4

Findings of the random effects model.
[DOCX File, 14 KB - [jmir_v24i3e34544_app4.docx](#)]

References

1. Labberton A, Godøy A, Elgersma I, Strand BH, Telle K, Arnesen T, et al. SARS-CoV-2 infections and hospitalisations among immigrants in Norway—significance of occupation, household crowding, education, household income and medical

- risk: a nationwide register study. *Scand J Public Health* 2022 Feb 14;14034948221075029 [FREE Full text] [doi: [10.1177/14034948221075029](https://doi.org/10.1177/14034948221075029)] [Medline: [35164616](https://pubmed.ncbi.nlm.nih.gov/35164616/)]
2. Indseth T, Grøslund M, Arnesen T, Skyrud K, Kløvstad H, Lamprini V, et al. COVID-19 among immigrants in Norway, notified infections, related hospitalizations and associated mortality: A register-based study. *Scand J Public Health* 2021 Feb;49(1):48-56 [FREE Full text] [doi: [10.1177/1403494820984026](https://doi.org/10.1177/1403494820984026)] [Medline: [33406993](https://pubmed.ncbi.nlm.nih.gov/33406993/)]
 3. Berg SH, O'Hara JK, Shortt MT, Thune H, Brønnick KK, Lungu DA, et al. Health authorities' health risk communication with the public during pandemics: a rapid scoping review. *BMC Public Health* 2021 Jul 15;21(1):1401 [FREE Full text] [doi: [10.1186/s12889-021-11468-3](https://doi.org/10.1186/s12889-021-11468-3)] [Medline: [34266403](https://pubmed.ncbi.nlm.nih.gov/34266403/)]
 4. Brekke J. Informing hard-to-reach immigrant groups about COVID-19—Reaching the Somali population in Oslo. *J Refug Stud* 2021;feab053. [doi: [10.1093/jrs/feab053](https://doi.org/10.1093/jrs/feab053)]
 5. Weekly reports for coronavirus and COVID-19. Norwegian Institute of Public Health. URL: <https://www.fhi.no/en/publ/2020/weekly-reports-for-coronavirus-og-covid-19/> [accessed 2021-09-08]
 6. Vinjerui KH, Elgersma IH, Fretheim A. Increased COVID-19 Testing Rates Following Combined Door-to-Door and Mobile Testing Facility Campaigns in Oslo, Norway, a Difference-in-Difference Analysis. *Int J Environ Res Public Health* 2021 Oct 21;18(21):11078 [FREE Full text] [doi: [10.3390/ijerph182111078](https://doi.org/10.3390/ijerph182111078)] [Medline: [34769597](https://pubmed.ncbi.nlm.nih.gov/34769597/)]
 7. Ipsos SoMe-tracker Q2'21 er endelig tilgjengelig. Ipsos. URL: <https://www.ipsos.com/nb-no/ipsos-some-tracker-q221-er-endelig-tilgjengelig> [accessed 2021-09-02]
 8. Create a blocked randomisation list. Sealed Envelope. URL: <https://www.sealedenvelope.com/simple-randomiser/v1/lists> [accessed 2021-09-10]
 9. So J, Kim S, Cohen H. Message fatigue: Conceptual definition, operationalization, and correlates. *Communication Monographs* 2016 Nov 02;84(1):5-29. [doi: [10.1080/03637751.2016.1250429](https://doi.org/10.1080/03637751.2016.1250429)]
 10. Kim S, So J. How Message Fatigue toward Health Messages Leads to Ineffective Persuasive Outcomes: Examining the Mediating Roles of Reactance and Inattention. *J Health Commun* 2018;23(1):109-116. [doi: [10.1080/10810730.2017.1414900](https://doi.org/10.1080/10810730.2017.1414900)] [Medline: [29272208](https://pubmed.ncbi.nlm.nih.gov/29272208/)]
 11. Emergency preparedness register for COVID-19 (Beredt C19). Norwegian Institute of Public Health. URL: <https://www.fhi.no/en/id/infectious-diseases/coronavirus/emergency-preparedness-register-for-covid-19/> [accessed 2021-09-09]
 12. Different ways to estimate treatment effects in randomised controlled trials. *Contemp Clin Trials Commun* 2018 Jun;10:80-85 [FREE Full text] [doi: [10.1016/j.conctc.2018.03.008](https://doi.org/10.1016/j.conctc.2018.03.008)] [Medline: [29696162](https://pubmed.ncbi.nlm.nih.gov/29696162/)]
 13. Barnett AG, van der Pols JC, Dobson AJ. Regression to the mean: what it is and how to deal with it. *Int J Epidemiol* 2005 Feb;34(1):215-220. [doi: [10.1093/ije/dyh299](https://doi.org/10.1093/ije/dyh299)] [Medline: [15333621](https://pubmed.ncbi.nlm.nih.gov/15333621/)]
 14. Potter W. *Media Effects*. Thousand Oaks, CA: Sage Publications; 2012.
 15. Alsan M, Stanford FC, Banerjee A, Breza E, Chandrasekhar AG, Eichmeyer S, et al. Comparison of Knowledge and Information-Seeking Behavior After General COVID-19 Public Health Messages and Messages Tailored for Black and Latinx Communities : A Randomized Controlled Trial. *Ann Intern Med* 2021 Apr;174(4):484-492 [FREE Full text] [doi: [10.7326/M20-6141](https://doi.org/10.7326/M20-6141)] [Medline: [33347320](https://pubmed.ncbi.nlm.nih.gov/33347320/)]
 16. Vandormael A, Adam M, Greuel M, Gates J, Favaretti C, Hachaturyan V, et al. The Effect of a Wordless, Animated, Social Media Video Intervention on COVID-19 Prevention: Online Randomized Controlled Trial. *JMIR Public Health Surveill* 2021 Jul 27;7(7):e29060 [FREE Full text] [doi: [10.2196/29060](https://doi.org/10.2196/29060)] [Medline: [34174778](https://pubmed.ncbi.nlm.nih.gov/34174778/)]
 17. Alatas V, Chandrasekhar A, Mobius M, Olken B, Paladines C. When Celebrities Speak: A Nationwide Twitter Experiment Promoting Vaccination In Indonesia. National Bureau of Economic Research. 2019. URL: https://www.nber.org/system/files/working_papers/w25589/w25589.pdf [accessed 2022-03-16]
 18. Breza E, Stanford FC, Alsan M, Alsan B, Banerjee A, Chandrasekhar AG, et al. Effects of a large-scale social media advertising campaign on holiday travel and COVID-19 infections: a cluster randomized controlled trial. *Nat Med* 2021 Sep;27(9):1622-1628 [FREE Full text] [doi: [10.1038/s41591-021-01487-3](https://doi.org/10.1038/s41591-021-01487-3)] [Medline: [34413518](https://pubmed.ncbi.nlm.nih.gov/34413518/)]
 19. Indseth T, Godøy A, Kjøllesdal M. Covid-19 etter fødeland fra mars 2020 til februar 2021. Folkehelseinstituttet. URL: <https://www.fhi.no/globalassets/dokumenterfiler/rapporter/2021/covid-19-etter-fodeland-fra-mars-2020-til-februar-2021-rapport-2021.pdf> [accessed 2021-09-02]

Abbreviations

- Beredt C19:** emergency preparedness register for COVID-19
MSIS: Norwegian Surveillance System for Communicable Diseases
NIPH: Norwegian Institute of Public Health
OR: odds ratio
PCR: polymerase chain reaction
RCT: randomized controlled trial

Edited by C Basch; submitted 03.11.21; peer-reviewed by I Mircheva, A Joshi, S Case; comments to author 11.12.21; revised version received 26.01.22; accepted 24.02.22; published 24.03.22.

Please cite as:

Elgersma IH, Fretheim A, Indseth T, Munch AT, Johannessen LB, Hansen CE

The Evaluation of a Social Media Campaign to Increase COVID-19 Testing in Migrant Groups: Cluster Randomized Trial

J Med Internet Res 2022;24(3):e34544

URL: <https://www.jmir.org/2022/3/e34544>

doi: [10.2196/34544](https://doi.org/10.2196/34544)

PMID: [35285811](https://pubmed.ncbi.nlm.nih.gov/35285811/)

©Ingeborg Hess Elgersma, Atle Fretheim, Thor Indseth, Anita Thorolvsen Munch, Live Bøe Johannessen, Christine Engh Hansen. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 24.03.2022. 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.

Original Paper

Tracking Public Attitudes Toward COVID-19 Vaccination on Tweets in Canada: Using Aspect-Based Sentiment Analysis

Hyeju Jang^{1,2}, MSc, PhD; Emily Rempel², MSc, PhD; Ian Roe², BComm, PBACC Com, MMIE; Prince Adu², MA, MPH, PhD; Giuseppe Carenini¹, MSc, PhD; Naveed Zafar Janjua^{2,3,4}, MBBS, MSc, DrPH

¹Department of Computer Science, University of British Columbia, Vancouver, BC, Canada

²British Columbia Centre for Disease Control, Vancouver, BC, Canada

³School of Population and Public Health, University of British Columbia, Vancouver, BC, Canada

⁴Centre for Health Evaluation and Outcome Sciences, St Paul's Hospital, Vancouver, BC, Canada

Corresponding Author:

Hyeju Jang, MSc, PhD

British Columbia Centre for Disease Control

655 West 12th Avenue

Vancouver, BC, V5Z 4R4

Canada

Phone: 1 778 580 9113

Email: hyejuj@cs.ubc.ca

Abstract

Background: The development and approval of COVID-19 vaccines have generated optimism for the end of the COVID-19 pandemic and a return to normalcy. However, vaccine hesitancy, often fueled by misinformation, poses a major barrier to achieving herd immunity.

Objective: We aim to investigate Twitter users' attitudes toward COVID-19 vaccination in Canada after vaccine rollout.

Methods: We applied a weakly supervised aspect-based sentiment analysis (ABSA) technique, which involves the human-in-the-loop system, on COVID-19 vaccination-related tweets in Canada. Automatically generated aspect and opinion terms were manually corrected by public health experts to ensure the accuracy of the terms and make them more domain-specific. Then, based on these manually corrected terms, the system inferred sentiments toward the aspects. We observed sentiments toward key aspects related to COVID-19 vaccination, and investigated how sentiments toward "vaccination" changed over time. In addition, we analyzed the most retweeted or liked tweets by observing most frequent nouns and sentiments toward key aspects.

Results: After applying the ABSA system, we obtained 170 aspect terms (eg, "immunity" and "pfizer") and 6775 opinion terms (eg, "trustworthy" for the positive sentiment and "jeopardize" for the negative sentiment). While manually verifying or editing these terms, our public health experts selected 20 key aspects related to COVID-19 vaccination for analysis. The sentiment analysis results for the 20 key aspects revealed negative sentiments related to "vaccine distribution," "side effects," "allergy," "reactions," and "anti-vaxxer," and positive sentiments related to "vaccine campaign," "vaccine candidates," and "immune response." These results indicate that the Twitter users express concerns about the safety of vaccines but still consider vaccines as the option to end the pandemic. In addition, compared to the sentiment of the remaining tweets, the most retweeted or liked tweets showed more positive sentiment overall toward key aspects ($P < .001$), especially vaccines ($P < .001$) and vaccination ($P = .009$). Further investigation of the most retweeted or liked tweets revealed two opposing trends in Twitter users who showed negative sentiments toward vaccines: the "anti-vaxxer" population that used negative sentiments as a means to discourage vaccination and the "Covid Zero" population that used negative sentiments to encourage vaccinations while critiquing the public health response.

Conclusions: Our study examined public sentiments toward COVID-19 vaccination on tweets over an extended period in Canada. Our findings could inform public health agencies to design and implement interventions to promote vaccination.

(*J Med Internet Res* 2022;24(3):e35016) doi:[10.2196/35016](https://doi.org/10.2196/35016)

KEYWORDS

COVID-19; vaccination; Twitter; aspect-based sentiment analysis; Canada; social media; pandemic; content analysis; vaccine rollout; sentiment analysis; public sentiment; public health; health promotion; vaccination promotion

Introduction

The development and approval of COVID-19 vaccines have generated optimism for the end of the COVID-19 pandemic and a return to normalcy. Current data from clinical trials and real-world studies show that these vaccines have high efficacy and effectiveness in preventing infection and severe disease [1-3]. Since the start of vaccination in Canada and many countries with high vaccine coverage, deaths among vaccinated groups have plummeted, and most new infections and hospitalizations are occurring among unvaccinated individuals, although recently more breakthrough infections have been reported with the Omicron variant. However, the overall impact on reduction in transmission, morbidity, mortality, as well as the easing of restrictions will depend on vaccine coverage and uptake. Vaccine acceptance and uptake remains one of the most important public health concerns in many countries. Vaccine hesitancy—often fueled by misinformation surrounding the importance, safety, or effectiveness of the vaccine—poses a major barrier to achieving herd immunity [4]. The failure to eradicate other infectious diseases including polio, for instance, has been blamed on conspiracy theories that fueled vaccine hesitancy [5]. The same concern is now threatening the successful implementation of mass COVID-19 vaccination campaigns. A systematic review of 126 surveys published before November 2020 showed an increasing worldwide hesitancy toward the COVID-19 vaccine [6].

In an era of misinformation and disinformation, identifying and understanding vaccine hesitancy are key to containing the spread of the virus and crucial to framing public health messaging [4]. In addition to individual views on the vaccine, negative discourse and misinformation related to it on social media could create doubt. Similar to infection, misinformation on social media spreads within the social network (echo chambers) and could affect acceptance among people. A recent study noted a decline in vaccine acceptance from exposure to misinformation on social media [7]. In addition to broader vaccine hesitancy, vaccine hesitancy in local areas or networks could also increase the risk of transmission within the community even if the broader population is immune. Tackling misinformation about the vaccine on social media will be critical to achieving herd immunity and a path out of the pandemic.

Various studies have investigated people's attitudes toward COVID-19 vaccination and addressed vaccine misinformation in social media. Puri et al [8] discussed digital health strategies to handle vaccine misinformation on social media such as

framing messages and leveraging celebrities. Chou and Budenz [9] reported that people's reactions to the COVID-19 vaccines are emotionally charged, and emphasized the role of emotion in vaccine communication efforts to address vaccine hesitancy. Thelwall et al [10] manually categorized 446 COVID-19 vaccine-hesitant tweets in English (March 10 to December 5, 2020) into 14 classes on the basis of what types of vaccine hesitancy information are shared. Some of these classes were efficacy, safety, and conspiracy. Hussain et al [11] performed sentiment analysis on Facebook and Twitter data to understand public sentiments toward COVID-19 vaccines in the United Kingdom and the United States. Guntuku et al [12] reported geographical and temporal variation in concerns about COVID-19 vaccines in the United States by using topic modeling. Kwok et al [13] analyzed tweets from Australian users by using topic modeling and sentiment analysis. Griffith et al [14] conducted content analysis on tweets from Canada, posted between December 10 and December 23, 2020, and identified reasons underlying vaccine hesitancy among Twitter users in Canada. Although prior research investigated public attitudes toward COVID-19 vaccination, to our knowledge, there has not been much research examining public sentiments toward COVID-19 vaccination over an extended period following the start of vaccine rollout. In addition, no prior work has investigated the most influential tweets.

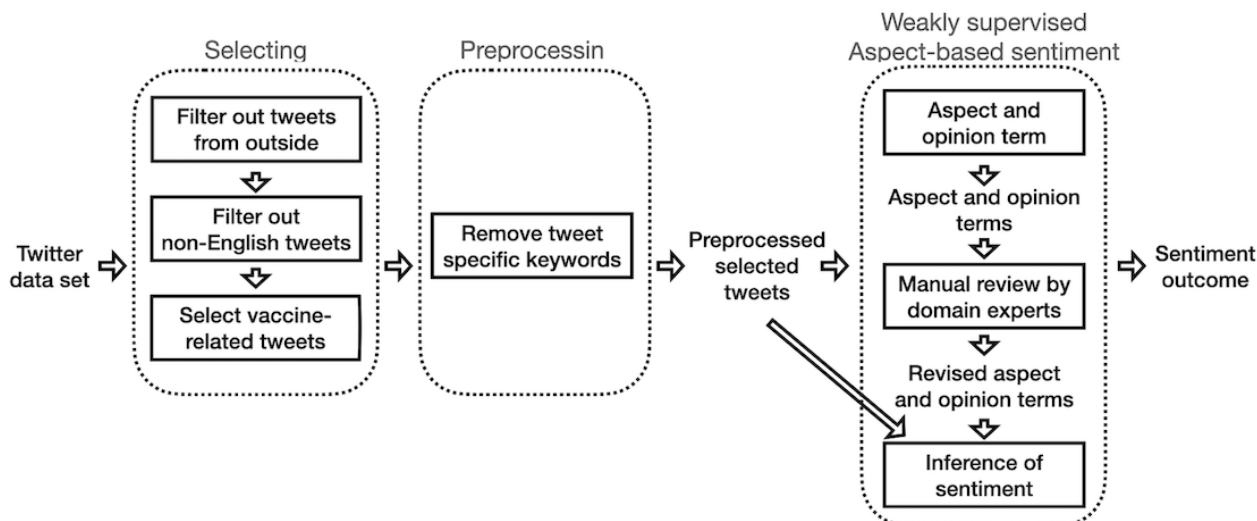
In this paper, we investigate Twitter users' attitudes toward COVID-19 vaccination in Canada from December 2020 to May 2021. More specifically, we examined sentiments toward certain aspects related to COVID-19 vaccination, which were chosen by our public health domain experts, and show how those sentiments change over time. In addition, we observe what the most retweeted or liked tweets discuss and how their sentiments are. For these analyses, we use weakly supervised aspect-based sentiment analysis (ABSA) [15], a natural language processing technique that allows identifying sentiments for aspects. Our findings could inform the design and implementation of public health interventions to promote vaccination, toward the goal of ending the pandemic.

Methods

Methods Overview

The overview of our process is displayed in [Figure 1](#). We first selected COVID-19 vaccination-related tweets and then preprocessed them. Then, based on these tweets, we performed weakly supervised ABSA.

Figure 1. Study overview.



Data and Preprocessing

We used a public Twitter data set about the COVID-19 pandemic, which was curated by Chen et al [16], using numerous COVID-19–related case-insensitive keywords such as “coronavirus,” “COVID-19,” and “pandemic.” Data collection started on January 28, 2020 (tweets from January 21, 2020), and is still ongoing. For the analysis in this paper, we collected tweets from December 2020 to May 2021, which spans from the beginning of vaccine rollout in Canada to the present.

From the data set, we extracted tweets originating in Canada and written in English by using tweet metadata and the spacy-langdetect toolkit (version 0.1.2) [17]. In addition, to focus on COVID-19 vaccination, we selected vaccine-related tweets using keywords (eg, “vaccine,” “vaccination,” “immunity,” “immune,” “rollout,” “mrna,” and “side effect”) defined by our domain experts. Our experts selected these keywords on the basis of public questions they received at the British Columbia Centre for Disease Control, which is responsible for vaccine programs and vaccine promotion. The original COVID-19 data set collected by Chen et al [16] contains 678,456,379 tweets for December 2020 to May 2021. Among them, the number of English tweets in Canada was 193,071. Of these, the number of the tweets that contain our search keywords was 21,821. As a result, 21,821 tweets were used for analysis. Among these, 10,202 tweets were original tweets, 6664 tweets were retweets or quoted tweets, 4955 tweets were replies to other tweets, and 112 tweets were both replies and quoted tweets. Of note, a quoted tweet is a type of retweet that contains an additional comment of the user. To remove tweet-specific keywords and URLs, we used the tweet-preprocessor toolkit (version 0.5.0) [18]. We did not remove hashtags and mentions because they are often part of a sentence, and could be informative for our work.

ABSA

We used ABSApp [19], a weakly supervised ABSA human-in-the-loop system, which enables public health domain experts to modify aspect and opinion terms that were automatically extracted from the corpus by the system. More

specifically, ABSApp automatically extracts both aspect and opinion terms from the data set without any manual labeling as the initial step and then allows humans to revise these terms. Then, based on these terms, the system automatically infers sentiment appearing in the data toward the aspect terms (Figure 1). In comparison to this weakly supervised ABSA system, a supervised method uses human labeled data and an unsupervised method does not use human efforts at all.

The aspect terms represent aspects we want to compute sentiment for, such as “pfizer” or “side effects.” The opinion terms are those that can be used for inference of sentiment; for example, “sore” for a negative sentiment and “immune” for a positive sentiment. The aspect and opinion terms can be domain-specific; hence, it is important to allow domain experts to edit them rather than just using automatically extracted general terms. These modified aspect and opinion terms were used for inference of sentiments for aspects.

To extract aspect and opinion terms in an unsupervised way using the ABSApp system first, we used 9239 tweets from December 2020 to February 2021. We used only the subset of the data for computational efficiency. The original study by Pereg et al [19] extracted terms from 75% of 2 data sets that consist of 5841 and 3614 sentences, respectively, which were much smaller than our data used for term extraction. Previous work [20] showed experimental results suggesting that increasing amounts of data beyond a certain point are unlikely to significantly increase the number of aspect terms when using ABSApp. In addition, opinion terms are not expected to differ much between February 2021 and May 2021 as they are the terms that indicate the sentiment of the tweets. The system automatically extracted 108 aspect terms and 6793 opinion terms.

Automatically extracted terms were manually reviewed and edited by 2 public health experts. Each expert reviewed and edited the whole list of aspect terms and opinion terms separately. The discrepancies in edits were addressed differently for the aspect terms and the opinion terms. For the aspect terms, we did not discard differences and selected all terms including discrepancies because an aspect term could be meaningful if at

least one expert found it useful. For the opinion terms, we excluded the terms that did not agree because those opinion terms would not be a strong indicator of a sentiment if the experts did not agree. This manual edition process resulted in 170 aspect terms and 6775 opinion terms, and they were used for the sentiment inference step. Among the aspect terms, our experts selected 20 key aspects more relevant to COVID-19 vaccination, for which we report results in this paper.

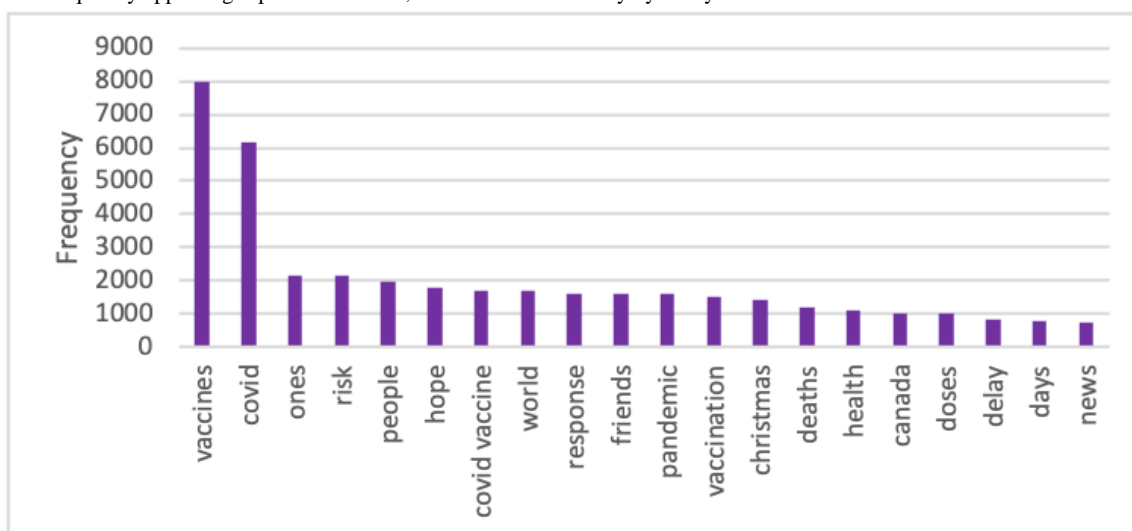
In addition, we investigated most liked or retweeted tweets to determine what they discuss and which sentiment they convey. We first identified original tweets that were not retweets, quoted tweets, or replies to other tweets. Tweet metadata such as *retweeted_status*, *is_quote_status*, and *in_reply_to_status_id* were used. Then, we sorted the original tweets by their *retweet_count* and *favorite_count* (Twitter changed the “favorite” button to a “like” button; hence, hereinafter, we have used these two terms interchangeably). To observe the discourse in the top-ranked original tweets, we computed the frequency

of nouns appearing in the tweets. For this computation, we first tokenized preprocessed tweets that did not contain tweet-specific keywords, URLs, and hashtags, and tagged part-of-speech. For both tokenization and part-of-speech tagging, the spacy toolkit (version 2.1.8) [21] was used. Scripts for preprocessing and analysis are available on the internet [22].

Results

Figure 2 shows the most frequently appearing aspect terms that were automatically extracted by the system. “vaccines” and “covid” were the most common aspects, probably because they were used for filtering out non-vaccine-related tweets during the data collection. The other words describe the discussion in this data set. People discussed aspects such as “hope,” “people,” “friends,” “world,” “deaths,” and “delay” together with COVID-19 vaccines. The less meaningful aspects (eg, “ones”) are included as these aspects were automatically extracted without any manual intervention.

Figure 2. Most frequently appearing aspects in the data, extracted automatically by the system.



The sentiment outcomes of the aspects most relevant to COVID-19 vaccines, which were selected by public health experts, are shown in Figure 3. Overall, the sentiment toward “vaccines,” “vaccination,” and specific vaccine brands such as “pfizer” and “moderna” show similar levels of positive and negative sentiments. A more negative sentiment was observed with “vaccine distribution,” “side effects,” “allergy,” “reactions,” and “anti-vaxxer.” A more positive sentiment was observed toward “vaccination campaign,” “vaccine candidates,” and “immune response.”

Figure 4 shows the change in sentiment over time toward “vaccination.” First, we observed that the discussion about vaccination increased drastically in March and peaked in mid-April. Thereafter, it decreased.

Tables 1 and 2 show the distribution of how many times the original tweets were retweeted and liked by other users,

respectively. As expected, large number of tweets were not retweeted or liked, while a small number of tweets had a large number of retweets and likes.

Figure 5 displays nouns that appeared in more than 5 tweets among the 100 most retweeted original tweets. These tweets were the same as the 100 most liked original tweets. In addition, Table 3 lists top 3 most liked and retweeted tweets. These data show that the most influential tweets were related to vaccine rollout, distribution, or administration and people’s hope and wish to end the pandemic situation.

Figure 6 shows the sentiment of 500 most liked tweets toward the key aspects selected by our public health experts. Compared to the sentiment of all the remaining tweets, these tweets show a significantly more positive sentiment overall about the key aspects ($P < .001$), especially vaccines ($P < .001$) and vaccination ($P = .009$), as revealed through chi-square analysis.

Figure 3. Results of aspect-based sentiment analysis. x-axis: selected aspects, y-axis (left): number of positive occurrences and number of negative occurrences in log scale; secondary y-axis (right): percentage of positive occurrences.

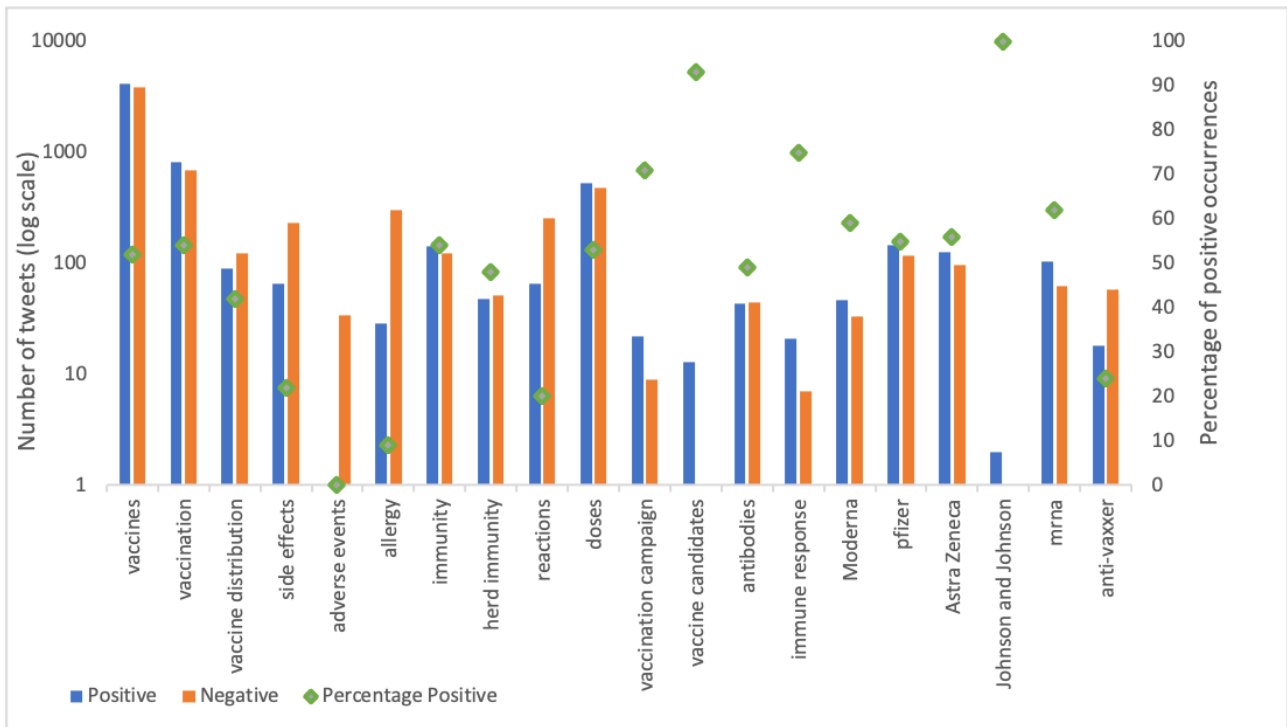


Figure 4. Sentiment changes over time for COVID-19 vaccination. Primary y-axis (left): number of total occurrences; secondary y-axis (right): percentage of positive occurrences.

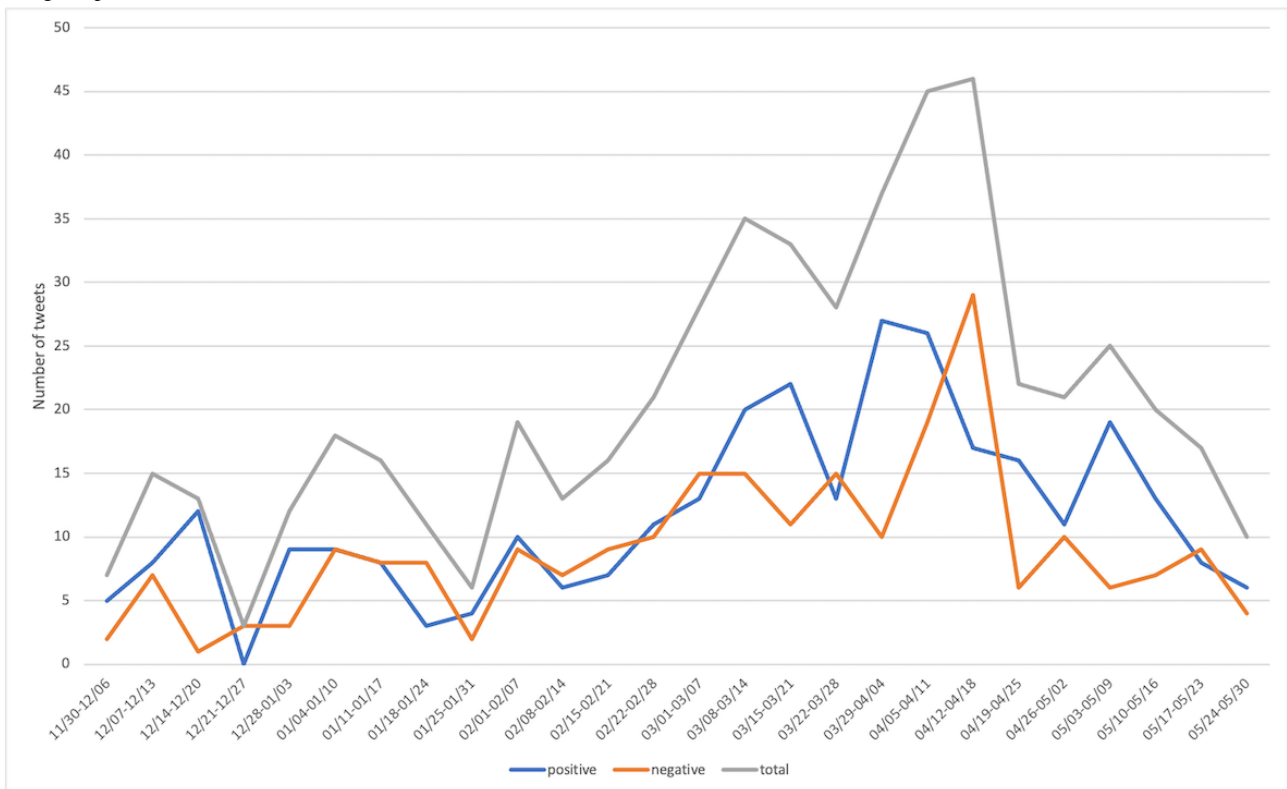


Table 1. Frequency distribution of original tweets being retweeted.

Number of times a tweet is retweeted	Frequency (number of tweets)
0	7033
1-50	3040
51-100	50
101-200	46
201-2169	33

Table 2. Frequency distribution of original tweets being liked.

Number of times a tweet is liked	Frequency (number of tweets)
0	4804
1-50	4863
51-100	238
101-200	126
201-400	79
401-800	52
801-7304	40

Figure 5. Terms that occur in more than 5 documents among the 100 most retweeted original tweets.

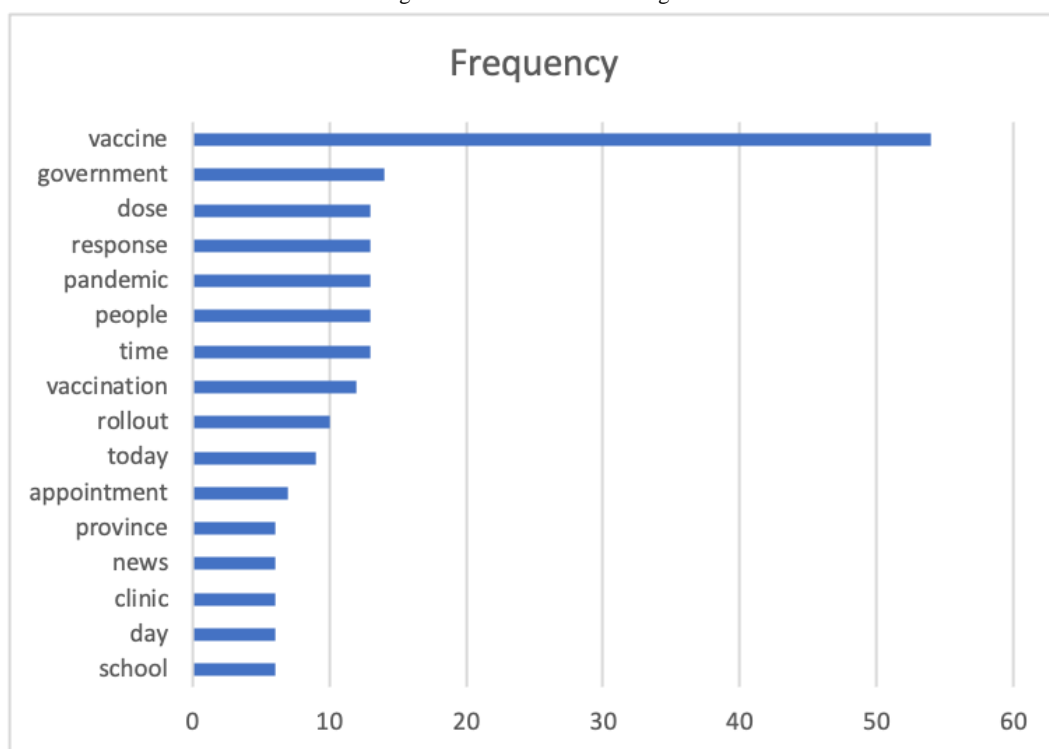
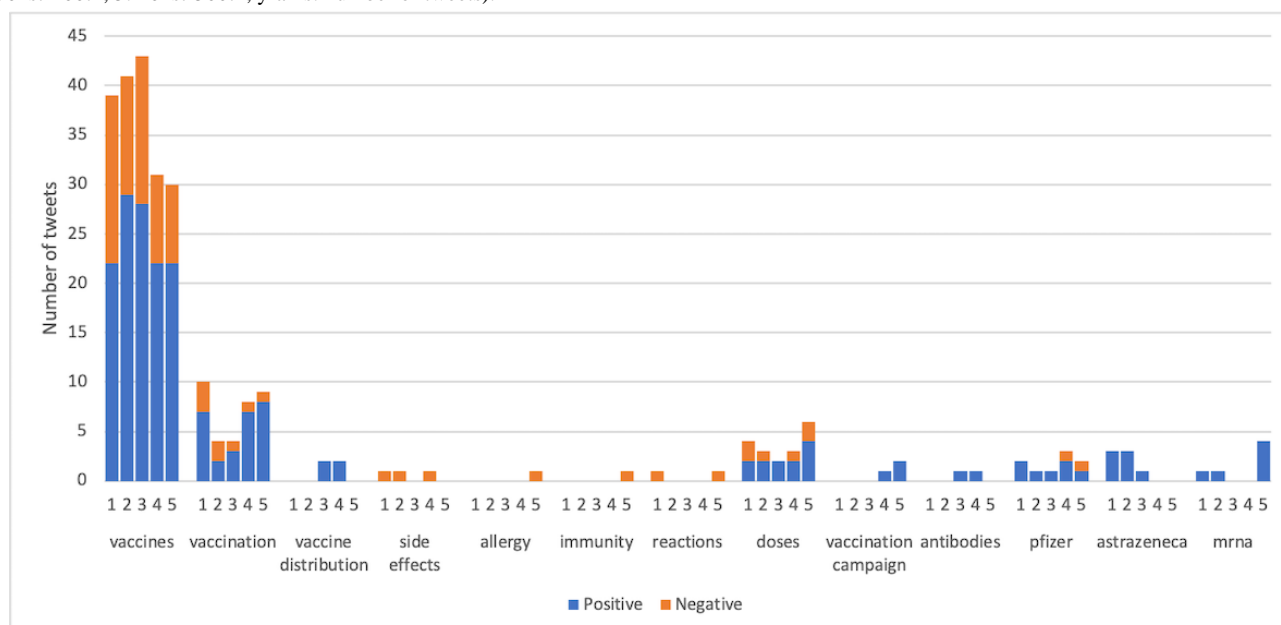


Table 3. Top 3 most liked original tweets (identifying information was removed).

Tweets	Likes, n	Retweets, n	Followers of the author, n
Do you know why they haven't lifted the pandemic "emergency" yet? The first sentence on this pfizer website has an answer: none of the vaccines have been approved. They are only authorized for use in an "emergency". No emergency, no experimental vaccines.	4306	2169	256,484
We have a hyperinfectious Covid variant circulating in the community, a vaccine shortage and are just on the verge of restoring contact tracing. If we pursue a #COVIDzero strategy we can be out of this by March. If we don't we can forfeit our sacrifices & see a huge 3rd wave.	4169	1329	20,352
Vaccines sitting in freezers in Ontario: 985,132 Total vaccines administered today: 59,567	2926	1061	23,923
please stop playing games with our lives. Implement #PaidSickDays, ramp up vaccine administrations, and a real lockdown. Mockdown 3.0 isn't cutting it.			

Figure 6. Sentiment toward certain aspects observed in the 500 most influential (most liked) tweets (1: top 100 tweets, 2: 101st-200th, 3: 201st-300th, 4: 301st-400th, 5: 401st-500th, x-axis: number of tweets).



Discussion

Principal Findings

In this study, we examined sentiments expressed toward COVID-19 vaccination on Twitter in Canada from December 2020 to May 2021. Major aspects included risk, hope, deaths, and delay. Positive sentiments were expressed for “vaccination campaign,” “vaccine candidates,” and “immune response,” while negative sentiments were expressed for issues related vaccine distribution, adverse effects (safety), and antivaccination. As expected, there was a small number of tweets that were retweeted multiple times. Within this group, we found antivaccination tweets, which were retweeted multiple times to a large number of people. Tackling misinformation spread by this small number of posts could have a large impact on countering vaccine misinformation.

The aspect “risk” in our most frequent aspects (Figure 2) suggests that people who were concerned about the risk of COVID-19 considered vaccination a solution. There were also tweets about the risk of side effects of vaccination, but mostly

the conversations were centered around the risk of COVID-19 outweighing the risk of vaccination. This is aligned with prior work, which revealed that the safety category was ranked in the top 3 among 446 COVID-19 vaccine–hesitant tweets between March 10 and December 5, 2020 [10]. Survey data on vaccine acceptance also indicate an association between concerns about vaccine safety and a low intention to receive a COVID-19 vaccine [23–25]. A similar finding was observed in the study by Lyu et al [26], which showed that “trust” was the most dominant emotion in tweets from March 11, 2020, to January 31, 2021. This highlights the need to address safety concerns about vaccines to increase vaccine uptake. Other frequently discussed aspects “delay” and “hope” in our data indicate that people were worried about the delay in vaccine rollout in Canada but still considered vaccines as the option to end the pandemic. These aspects highlight frustration from twitter users in the delayed vaccine supply in Canada compared to those in the United States and in other countries.

Our results show that the largest segment of negative sentiment was toward vaccine adverse effects including “allergy,” “side

effects,” and “reactions” (Figure 3). This finding is similar to prior work that showed that the largest public concern about COVID-19 vaccines was safety. Eibensteiner et al [27] reported that 41.7% of Twitter users (n=3439) around the world were unsure about the safety of COVID-19 vaccines, in the polls conducted for 1 week in mid-February 2021. In addition, in the earlier analysis performed on manually coded tweets in Canada on December 18 and 23, 2020, a total of 48.3% (292/605) of tweets expressed concerns about the safety of vaccines [24,25]. These concerns are consistent with the public health focus as well as media attention to side effects including thrombosis and myocarditis (heart inflammation). This indicates the continued challenge for public health agencies to address concerns about vaccine safety and instill confidence about COVID-19 vaccines.

It is also interesting to note the lack of difference in sentiment toward the vaccine developed by pfizer versus that developed by AstraZeneca (Figure 3). This might suggest a more complex understanding of vaccination by Twitter users despite media attention toward people not wanting to be vaccinated with AstraZeneca’s vaccine in March 2021 [28,29]. The high attentiveness to dosing intervals after vaccination (Figure 3) further emphasizes the challenges of knowledge translation and science communication as evidence changes.

Our results (Figure 4) show that the discussion about vaccination gradually increased after vaccine rollout on December 14, 2020, and drastically decreased after mid-April 2021. This finding aligns with the deceleration in the uptake of the first vaccine dose [30,31]. This may indicate a shift of public attention away from vaccination and the need to bring back a focus on vaccination to enhance vaccine coverage.

The 500 most liked tweets showed more positive sentiments overall toward the key aspects outlined in Figure 6, especially toward vaccines and vaccination, compared to the sentiments from the remaining tweets. On closer inspection, the most liked or retweeted tweets (Table 3) showed an interesting dichotomy in Twitter users who reported a negative sentiment toward vaccines: the “anti-vaxxer” population that largely used negative sentiment as a means to discourage vaccination, and the “Covid Zero” population that also used negative sentiment, albeit to encourage vaccination, while also critiquing the public health response. This finding from the most influential tweets highlights the need for public health agencies to have a

multipronged approach to health messaging related to vaccine information and misinformation. In addition, given that the most liked or retweeted tweets have a very large circle of influence (eg, 256,484 followers), investigating these tweets is crucial to identifying and tackling misinformation and even tracking dissemination of positive information. For example, if such tweets could be flagged for public health agencies through a dashboard, the information or claim in them could be verified or refuted in official channels.

Limitations

This study highlights important trends in Twitter conversations, which may have crucial policy implications. However, the following limitations need to be considered when interpreting the results and when our pipeline is deployed in practice. First, the tweets we used might not be representative of the general population. Recently, it is reported that the largest age group in Twitter users is 30-49 years (44%) [32]. Second, it should be noted that specific cities might overrepresent Canada in our study according to previous work by Gore et al [33]. However, the Twitter data that meet our inclusion criteria are not sufficient for a city-level analysis, especially given that we carry out our analysis on the basis of aspects. The limitations we mention here exist not only for our study but also for most previous work, which used tweets for a certain period at a nationwide level.

Second, although the ABSApp approach for ABSA leverages manually edited domain-specific aspect and opinion terms to improve the accuracy of sentiment analysis, it is still premature to accurately capture the intricacy of human language, such as figurative languages. Nevertheless, these limitations are not particular to our study but rather inherent in all current state-of-the-art sentiment analysis tools and techniques as well.

Conclusions

Social media presents a convenient opportunity to assess public perception on a wide variety of issues, as part of the pandemic response. This study examined public sentiments toward COVID-19 vaccination on tweets over an extended period in Canada. In addition, we also investigated the most influential tweets. Our findings could inform public health agencies to design and implement interventions to promote vaccination, and approach the goal of ending the pandemic.

Authors' Contributions

HJ, ER, IR, GC, and NZJ contributed to the conception and design of the study. HJ performed data collection, preprocessing, and all analyses. ER and IR contributed to reviewing automatically generated aspect and opinion terms. GC, PA, and NZJ contributed to the discussion of statistical analysis. HJ wrote the first draft of the manuscript. ER contributed to sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Conflicts of Interest

None declared.

References

1. De-Leon H, Calderon-Margalit R, Pederiva F, Ashkenazy Y, Gazit D. First indication of the effect of COVID-19 vaccinations on the course of the outbreak in Israel. medRxiv. Preprint posted online February 3, 2021 . [doi: [10.1101/2021.02.02.21250630](https://doi.org/10.1101/2021.02.02.21250630)]

2. Rossman H, Shilo S, Meir T, Gorfine M, Shalit U, Segal E. COVID-19 dynamics after a national immunization program in Israel. *Nat Med* 2021 Jun;27(6):1055-1061. [doi: [10.1038/s41591-021-01337-2](https://doi.org/10.1038/s41591-021-01337-2)] [Medline: [33875890](https://pubmed.ncbi.nlm.nih.gov/33875890/)]
3. Christie A, Henley SJ, Mattocks L, Fernando R, Lansky A, Ahmad FB, et al. Decreases in COVID-19 Cases, Emergency Department Visits, Hospital Admissions, and Deaths Among Older Adults Following the Introduction of COVID-19 Vaccine - United States, September 6, 2020-May 1, 2021. *MMWR Morb Mortal Wkly Rep* 2021 Jun 11;70(23):858-864 [FREE Full text] [doi: [10.15585/mmwr.mm7023e2](https://doi.org/10.15585/mmwr.mm7023e2)] [Medline: [34111059](https://pubmed.ncbi.nlm.nih.gov/34111059/)]
4. Sallam M. COVID-19 Vaccine Hesitancy Worldwide: A Concise Systematic Review of Vaccine Acceptance Rates. *Vaccines (Basel)* 2021 Feb 16;9(2):160 [FREE Full text] [doi: [10.3390/vaccines9020160](https://doi.org/10.3390/vaccines9020160)] [Medline: [33669441](https://pubmed.ncbi.nlm.nih.gov/33669441/)]
5. Khan Y, Mallhi T, Alotaibi N, Alzarea A, Alanazi A, Tanveer N, et al. Threat of COVID-19 Vaccine Hesitancy in Pakistan: The Need for Measures to Neutralize Misleading Narratives. *Am J Trop Med Hyg* 2020 Aug;103(2):603-604 [FREE Full text] [doi: [10.4269/ajtmh.20-0654](https://doi.org/10.4269/ajtmh.20-0654)] [Medline: [32588810](https://pubmed.ncbi.nlm.nih.gov/32588810/)]
6. Schwarzing M, Watson V, Arwidson P, Alla F, Luchini S. COVID-19 vaccine hesitancy in a representative working-age population in France: a survey experiment based on vaccine characteristics. *Lancet Public Health* 2021 Apr;6(4):e210-e221 [FREE Full text] [doi: [10.1016/S2468-2667\(21\)00012-8](https://doi.org/10.1016/S2468-2667(21)00012-8)] [Medline: [33556325](https://pubmed.ncbi.nlm.nih.gov/33556325/)]
7. 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;5(3):337-348. [doi: [10.1038/s41562-021-01056-1](https://doi.org/10.1038/s41562-021-01056-1)] [Medline: [33547453](https://pubmed.ncbi.nlm.nih.gov/33547453/)]
8. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. *Hum Vaccin Immunother* 2020 Nov 01;16(11):2586-2593 [FREE Full text] [doi: [10.1080/21645515.2020.1780846](https://doi.org/10.1080/21645515.2020.1780846)] [Medline: [32693678](https://pubmed.ncbi.nlm.nih.gov/32693678/)]
9. 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](https://doi.org/10.1080/10410236.2020.1838096)] [Medline: [33124475](https://pubmed.ncbi.nlm.nih.gov/33124475/)]
10. Thelwall M, Kousha K, Thelwall S. Covid-19 vaccine hesitancy on English-language Twitter. *EPI* 2021 Mar 15. [doi: [10.3145/epi.2021.mar.12](https://doi.org/10.3145/epi.2021.mar.12)]
11. Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study. *J Med Internet Res* 2021 Apr 05;23(4):e26627 [FREE Full text] [doi: [10.2196/26627](https://doi.org/10.2196/26627)] [Medline: [33724919](https://pubmed.ncbi.nlm.nih.gov/33724919/)]
12. Guntuku SC, Bottenheim AM, Sherman G, Merchant RM. Twitter discourse reveals geographical and temporal variation in concerns about COVID-19 vaccines in the United States. *Vaccine* 2021 Jul 05;39(30):4034-4038 [FREE Full text] [doi: [10.1016/j.vaccine.2021.06.014](https://doi.org/10.1016/j.vaccine.2021.06.014)] [Medline: [34140171](https://pubmed.ncbi.nlm.nih.gov/34140171/)]
13. 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 May 19;23(5):e26953 [FREE Full text] [doi: [10.2196/26953](https://doi.org/10.2196/26953)] [Medline: [33886492](https://pubmed.ncbi.nlm.nih.gov/33886492/)]
14. Griffith J, Marani H, Monkman H. COVID-19 Vaccine Hesitancy in Canada: Content Analysis of Tweets Using the Theoretical Domains Framework. *J Med Internet Res* 2021 Apr 13;23(4):e26874 [FREE Full text] [doi: [10.2196/26874](https://doi.org/10.2196/26874)] [Medline: [33769946](https://pubmed.ncbi.nlm.nih.gov/33769946/)]
15. Liu B. Sentiment Analysis and Subjectivity. In: Indurkha N, Damerau FJ, editors. *Handbook of Natural Language Processing* (2nd edition). Boca Raton, FL: CRC Press; 2010:627-666.
16. 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](https://doi.org/10.2196/19273)] [Medline: [32427106](https://pubmed.ncbi.nlm.nih.gov/32427106/)]
17. spacy-langdetect 0.1.2. Python Software Foundation. URL: <https://pypi.org/project/spacy-langdetect/> [accessed 2021-11-16]
18. tweet-preprocessor 0.6.0. Python Software Foundation. URL: <https://pypi.org/project/tweet-preprocessor/> [accessed 2021-11-16]
19. Pereg O, Korat D, Wasserblat M, Mamou J, Dagan I. ABSApp: A Portable Weakly-Supervised Aspect-Based Sentiment Extraction System. 2019 Presented at: 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations; 2019; Hong Kong. [doi: [10.18653/v1/d19-3001](https://doi.org/10.18653/v1/d19-3001)]
20. Johnson D, Dragojlovic N, Kopac N, Chen Y, Lenzen M, Le HS. EXPECT-NLP: An Integrated Pipeline and User Interface for Exploring Patient Preferences Directly from Patient-Generated Text. *Studies in Computational Intelligence* 2021.
21. Industrial-Strength Natural Language Processing. spaCy. URL: <https://spacy.io/> [accessed 2021-11-16]
22. hyejuj / covid19_vaccination. GitHub. URL: https://github.com/hyejuj/covid19_vaccination [accessed 2022-03-15]
23. Fisher KA, Bloomstone SJ, Walder J, Crawford S, Fouayzi H, Mazor KM. Attitudes Toward a Potential SARS-CoV-2 Vaccine : A Survey of U.S. Adults. *Ann Intern Med* 2020 Dec 15;173(12):964-973 [FREE Full text] [doi: [10.7326/M20-3569](https://doi.org/10.7326/M20-3569)] [Medline: [32886525](https://pubmed.ncbi.nlm.nih.gov/32886525/)]
24. Paul E, Steptoe A, Fancourt D. Attitudes towards vaccines and intention to vaccinate against COVID-19: Implications for public health communications. *Lancet Reg Health Eur* 2021 Feb;1:100012 [FREE Full text] [doi: [10.1016/j.lanpe.2020.100012](https://doi.org/10.1016/j.lanpe.2020.100012)] [Medline: [33954296](https://pubmed.ncbi.nlm.nih.gov/33954296/)]

25. Solís Arce JS, Warren SS, Meriggi NF, Scacco A, McMurry N, Voors M, et al. COVID-19 vaccine acceptance and hesitancy in low- and middle-income countries. *Nat Med* 2021 Aug;27(8):1385-1394 [FREE Full text] [doi: [10.1038/s41591-021-01454-y](https://doi.org/10.1038/s41591-021-01454-y)] [Medline: [34272499](https://pubmed.ncbi.nlm.nih.gov/34272499/)]
26. Lyu JC, Han EL, Luli GK. COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. *J Med Internet Res* 2021 Jun 29;23(6):e24435 [FREE Full text] [doi: [10.2196/24435](https://doi.org/10.2196/24435)] [Medline: [34115608](https://pubmed.ncbi.nlm.nih.gov/34115608/)]
27. Eibensteiner F, Ritschl V, Nawaz FA, Fazel SS, Tsagkaris C, Kulnik ST, et al. People's Willingness to Vaccinate Against COVID-19 Despite Their Safety Concerns: Twitter Poll Analysis. *J Med Internet Res* 2021 Apr 29;23(4):e28973 [FREE Full text] [doi: [10.2196/28973](https://doi.org/10.2196/28973)] [Medline: [33872185](https://pubmed.ncbi.nlm.nih.gov/33872185/)]
28. Modjeski M. Regina resident who thought she'd be given Pfizer vaccine shocked after being told she'd get AstraZeneca. *CBC News*. CBC News. 2021 Apr 10. URL: <https://www.cbc.ca/news/canada/saskatoon/vaccine-frustrations-astrazeneca-pfizer-1.5982861> [accessed 2021-08-13]
29. Lofaro J. Some Montrealers refusing AstraZeneca vaccine; Canadian authorities assert it's safe. *CTV News*. 2021 Mar 15. URL: <https://montreal.ctvnews.ca/some-montrealers-refusing-astrazeneca-vaccine-canadian-authorities-assert-it-s-safe-1.5347326> [accessed 2021-08-13]
30. COVID-19 vaccination in Canada. Government of Canada. URL: <https://health-infobase.canada.ca/covid-19/vaccination-coverage/> [accessed 2021-07-12]
31. Lambert S. After initial rush of keeners, uptake for COVID-19 vaccines on the Prairies has been slow growing. *CBC News*. 2021 Jun 17. URL: <https://www.cbc.ca/news/canada/edmonton/after-initial-rush-of-keeners-uptake-for-covid-19-vaccines-on-the-prairies-has-been-slow-growing-1.6069570> [accessed 2021-07-27]
32. Barnhart B. Social media demographics to inform your brand's strategy in 2022. *Sprout Social*. URL: <https://sproutsocial.com/insights/new-social-media-demographics/> [accessed 2021-10-27]
33. 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](https://doi.org/10.1371/journal.pone.0133505)] [Medline: [26332588](https://pubmed.ncbi.nlm.nih.gov/26332588/)]

Abbreviations

ABSA: aspect-based sentiment analysis

Edited by C Basch; submitted 16.11.21; peer-reviewed by L Huangfu, R Gore; comments to author 27.12.21; revised version received 22.02.22; accepted 22.02.22; published 29.03.22.

Please cite as:

Jang H, Rempel E, Roe I, Adu P, Carenini G, Janjua NZ

Tracking Public Attitudes Toward COVID-19 Vaccination on Tweets in Canada: Using Aspect-Based Sentiment Analysis

J Med Internet Res 2022;24(3):e35016

URL: <https://www.jmir.org/2022/3/e35016>

doi: [10.2196/35016](https://doi.org/10.2196/35016)

PMID: [35275835](https://pubmed.ncbi.nlm.nih.gov/35275835/)

©Hyeju Jang, Emily Rempel, Ian Roe, Prince Adu, Giuseppe Carenini, Naveed Zafar Janjua. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 29.03.2022. 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.

Original Paper

The Challenge of Debunking Health Misinformation in Dynamic Social Media Conversations: Online Randomized Study of Public Masking During COVID-19

Mehdi Mourali¹, PhD; Carly Drake², PhD

¹Haskayne School of Business, University of Calgary, Calgary, AB, Canada

²Department of Management and Marketing, North Central College, Naperville, IL, United States

Corresponding Author:

Mehdi Mourali, PhD

Haskayne School of Business

University of Calgary

2500 University Drive NW

Calgary, AB, T2N 1N4

Canada

Phone: 1 403 220 6684

Email: mehdi.mourali@haskayne.ucalgary.ca

Abstract

Background: The spread of false and misleading health information on social media can cause individual and social harm. Research on debunking has shown that properly designed corrections can mitigate the impact of misinformation, but little is known about the impact of correction in the context of prolonged social media debates. For example, when a social media user takes to Facebook to make a false claim about a health-related practice and a health expert subsequently refutes the claim, the conversation rarely ends there. Often, the social media user proceeds by rebuking the critic and doubling down on the claim.

Objective: The aim of this study was to examine the impact of such extended back and forth between false claims and debunking attempts on observers' dispositions toward behavior that science favors. We tested competing predictions about the effect of extended exposure on people's attitudes and intentions toward masking in public during the early days of the COVID-19 pandemic and explored several psychological processes potentially underlying this effect.

Methods: A total of 500 US residents took part in an online experiment in October 2020. They reported on their attitudes and intentions toward wearing masks in public. They were then randomly assigned to one of four social media exposure conditions (misinformation only vs misinformation+correction vs misinformation+correction+rebuke vs misinformation+correction+rebuke+second correction), and reported their attitudes and intentions for a second time. They also indicated whether they would consider sharing the thread if they were to see it on social media and answered questions on potential mediators and covariates.

Results: Exposure to misinformation had a negative impact on attitudes and intentions toward masking ($\beta=-.35$, 95% CI $-.42$ to $-.29$; $P<.001$). Moreover, initial debunking of a false claim generally improved attitudes and intentions toward masking ($\beta=.35$, 95% CI $.16$ to $.54$; $P<.001$). However, this improvement was washed out by further exposure to false claims and debunking attempts ($\beta=-.53$, 95% CI $-.72$ to $-.34$; $P<.001$). The latter result is partially explained by a decrease in the perceived objectivity of truth. That is, extended exposure to false claims and debunking attempts appear to weaken the belief that there is an objectively correct answer to how people ought to behave in this situation, which in turn leads to less positive reactions toward masking as the prescribed behavior.

Conclusions: Health professionals and science advocates face an underappreciated challenge in attempting to debunk misinformation on social media. Although engaging in extended debates with science deniers and other purveyors of bunk appears necessary, more research is needed to address the unintended consequences of such engagement.

(*J Med Internet Res* 2022;24(3):e34831) doi:[10.2196/34831](https://doi.org/10.2196/34831)

KEYWORDS

misinformation; debunking; correction; social media; truth objectivity; COVID-19; infodemiology; health information; digital health; public health; health professional

Introduction

Context

The internet in general and social media in particular have become important sources of information for many people seeking medical and health-related information. As of 2014, 72% of internet users in the United States reported having searched for health-related information online [1]. More recently, 49% of US adults reported obtaining at least some of their news about the COVID-19 vaccine on social media, and among those who regularly obtain news from social media, 61% rated social media as an important way of keeping up with news about COVID-19 vaccines [2]. Yet, content disseminated through social media sites such as Facebook, Twitter, or Reddit remains largely unregulated, and is replete with false and misleading information [3-8].

The widespread availability and consumption of false, inaccurate, or incomplete health information (herein referred to as “health misinformation”) is a serious problem that can cause individual and social harm by promoting erroneous beliefs about health and illness, leading to detrimental behavior [9,10]. For example, the persistent circulation of unfounded claims linking vaccination to autism has convinced many parents not to immunize their children, which has resulted in a significant rise in vaccine-preventable diseases and death [11]. Believing misinformation about COVID-19 has been linked to lower adoption of health protective behaviors [12] and greater consumption of harmful products [13]. Evidence also suggests that online disinformation campaigns on a global scale have played a key role in the notable drop in vaccination coverage over time [14].

Recognizing the potentially disastrous consequences of letting misinformation proliferate on social media, researchers from diverse fields have proposed a variety of countermeasures, including technological solutions aimed at limiting exposure to misinformation, educational interventions aimed at empowering people to recognize and deal with misinformation, as well as communication tools designed to help debunk and correct misinformation [15-19]. Importantly, health experts and health care professionals have been called upon to play an active role in correcting health misinformation when they encounter it on social media [20-27].

Elsewhere, a rich literature on debunking has shown that properly designed corrections can be effective at countering misinformation [28-31]. This research, however, has examined the issue mostly from a static perspective. In a typical study, participants are first exposed to misinformation. Subsequently, some participants receive a correction, and their responses are compared to a control group that received no correction or a comparison group that received an alternate correction varying in its content, source, or some other relevant attribute (eg, [27,32-36]). Although this paradigm allows for a clean test of the relative effectiveness of specific debunking interventions,

it oversimplifies the dynamic nature of social media conversations. For example, when a social media user takes to Facebook to make a false claim about a health-related practice, and a health expert subsequently refutes the claim, the conversation rarely ends there. Often, the social media user proceeds by rebuking the critic and doubling down on the claim.

Objectives

The aim of this study was to examine the impact of such extended back and forth between false claims and debunking attempts on observers’ dispositions toward behavior favored by science. The US Centers for Disease Control and Prevention (CDC) had been recommending mask wearing in public since April 2020. However, by the time of our study in October 2020, less than half of the states had issued a mandate for mask wearing in public [37], and misinformation regarding the safety and efficacy of masking had become rampant on social media [38].

In this study, we tested competing predictions about the effect of extended exposure on people’s attitudes and intentions toward masking in public and explored several psychological processes potentially underlying this effect.

On the one hand, properly debunking a false claim may have a lasting effect, such that when the purveyor of misinformation proceeds to rebuke the critic and double down on the false claim, observers’ attitudes toward masking would remain unaffected by the new round of misinformation. A detailed refutation that includes a clear explanation of why a claim is false and what is true instead [16,39] could have a persistent impact on observers’ attitudes because, in addition to arming them with facts, it undermines the credibility of the argument underlying the false claim [40]. Moreover, research on “prebunking” has shown that it is possible to inoculate individuals against misinformation before it is even encountered [41-43]. Extending the principle of prebunking to prolonged social media debates, one could surmise that witnessing a thorough refutation of a false claim early in the debate may inoculate people against later assertions of the false claim.

On the other hand, repeated exposure to misinformation and its rebuttal could create uncertainty about the very existence of true facts [4,44]. For example, in reference to President Donald Trump’s extensive record of making false and misleading claims, Lewandowski et al [44] quoted a 2017 editorial from the Bangor Daily News suggesting that one important consequence of the repeated falsehoods is that “A third of the population will say ‘clearly the White House is lying,’ a third will say ‘if Trump says it, it must be true,’ and the remaining third will say ‘gosh, I guess this is unknowable.’” A year later, former Director of National Intelligence, James Clapper [45], warned that “many Americans are questioning if facts are even knowable, as foreign adversaries and our national leaders continue to deny objective reality while advancing their own alternative facts.”

In the health domain, exposure to conflicting information about a wide range of topics, including mammography [46,47], nutrition [48,49], and the human papillomavirus vaccine [50], has been shown to increase confusion, uncertainty, and negative attitudes toward the health topic in question. Together, these results suggest that exposure to an extended back and forth between false claims and debunking attempts may generate doubt about the very existence of a nonsubjective true answer to a health-related question. This study specifically examined the discourse surrounding the question of whether wearing masks in public should be prescribed. Based on extant theory, we hypothesized that such discourse would weaken people's attitudes and intentions toward masking in public as the prescribed behavior.

In addition to testing these competing predictions, we explored the impact of extended exposure on people's intentions to share the social media threads. This has direct implications for understanding how misinformation spreads on social media.

Methods

Participants and Procedure

Respondents were recruited from Prolific, a platform for recruiting online participants that explicitly caters to researchers. Prolific provides diverse and valid samples, and its data quality compares favorably with that of other online platforms such as MTurk [51,52]. US residents with an existing account on Prolific took part in the study on October 16, 2020, in exchange for monetary compensation.

Upon consenting in writing, participants answered questions about the relevance of various sources of health information, the impact of COVID-19 (including their perceived risk of infection), and reported on their attitudes and intentions toward wearing masks in public. They then completed a first attention check and were randomly assigned to one of four social media exposure conditions (misinformation only [M] vs misinformation+correction [MC] vs misinformation+correction+rebuke [MCR] vs misinformation+correction+rebuke+second correction [MCRC]). The attention check consisted of a statement at the end of a question asking participants to choose a specific answer to ensure they were reading the questions.

The social media content consisted of Reddit posts that were adapted from real social media posts and expert responses collected by the International Fact Checking Network [38], which are shown in full in Figure 1. Participants in the M condition reviewed a post by a user with the username *citizen-health* arguing that wearing face masks can cause a new disease called "mask mouth." Those in the MC condition saw the same post plus a correction from a user with the username *Health_Scientist*, pointing out that there is in fact no scientific evidence of any new disease caused by wearing face masks. Those in the MCR group saw a thread containing the previous two posts plus a second post from *citizen-health* rebuking *Health_Scientist* and doubling down on the original claim. Finally, those in the MCRC condition saw a thread containing the previous three posts plus a second correction from *Health_Scientist*.

Figure 1. Reddit posts for the misinformation+correction+rebuttal+second correction condition.

u/citizen-health · Posted by u/citizen-health 43 minutes ago

Mask Mouth

"Mask Mouth" a new medical condition caused by wearing a masks all day. The bacteria in yr mouth along with the dampness causes the skin around your mouth to breakout in hives. Not to mention the oxygen deficiency due to the fact that you are breathing in your own carbon dioxide.

There is so much bacteria, yeast, virus & fungus that lives in the mouth that is expelled as you breath, but it is trapped around your mouth when wearing a mask all day. That along with the moisture from your breath creates the dampness that bacteria & fungus thrive in.

It's basically like wearing a soiled diaper all day. You will get a diaper rash.

Other mask mouth symptoms include tooth decay, bad breath & gum inflammation

10 Comments · Share · Save · ...

Health_Scientist · 33 minutes ago

The claims made by citizen-health are simply incorrect and unsupported. There is no evidence that wearing masks leads to bacterial overgrowth or causes infections. Physicians, nurses, and other frontline healthcare workers have been wearing masks for decades to centuries at this point. Some surgical procedures takes hours, sometimes 10, 12, or even 18 hours. Surgeons wear masks continuously during this period. There is no evidence that doing so increases risk of bacterial or fungal infection.

There is also no scientific basis for the claim that wearing masks causes dental problems like tooth decay, gum disease, or bad breath. Dentists themselves wear face masks for long periods of time without any impact on their oral health.


Also, face mask filters are small enough to keep out infectious droplets containing viruses, but gas molecules such as carbon dioxide and oxygen can still pass through freely. Healthcare workers who wear masks for long periods of time do not demonstrate significant impairment in work performance, as would be the case if masks did cause oxygen deficiency.

In some cases, prolonged wearing of face masks might cause skin irritation, but this can be prevented by selecting comfortable masks and following proper hygiene guidelines.

citizen-health · 15 minutes ago

Of course, a government shall like "Health_Scientist" would deny the existence of this new mouth mask. What else do you expect? Hey, they say "A picture is worth a thousand words" & this picture screams TRUTH 🤔🤔🤔

👉 TAKE YOUR NASTY GERM & BACTERIA INFESTED MASKS OFF! 🤢🤮🤨🤨🤨



Health_Scientist · just now

citizen-health is resorting to imagery to provoke an emotional response. But, scientific evidence does not support the claim that wearing masks increases the risk of oral and skin infections or bacterial overgrowth.

Masks are an incredibly effective public health tool, and essential in stemming the spread of SARS-CoV-2, the virus that causes COVID-19. While initially there was a lack of scientific certainty about the level of protection that different masks conferred, it is now clear that masks are helpful. And not just "my mask protects you, your mask protects me" (if I wear a mask, and am infected, I breath out and spread less virus), but beyond that, wearing a mask means that if you come into contact with the virus, you are likely to be exposed to a lower amount of virus, and give your body and immune system a chance to respond better, with better outcomes.

It is important to pay attention to where your information is coming from, and what it's based on. The recommendation to wear masks is from doctors, infectious disease experts, epidemiologists, and those leading the efforts to curb the pandemic. Countries with widespread mask use have lower rates of infection, fewer infected patients, and better outcomes - and are able to get back closer to 'normal' life, with their economy doing better, and more places of business and schools able to open, and able to stay open.

After reviewing the Reddit thread, participants completed a comprehension check assessing their comprehension of the position argued by *citizen-health*. They then indicated whether they would consider sharing the thread if they were to see it on social media, and reported their attitudes and intentions for a second time. They also answered questions on potential mediators and additional covariates. Mediators included perceptions of objectivity of truth, argument strength, and warmth and competence of the protagonists in the Reddit

exchange. Covariates included perceived COVID-19 risk, cognitive reflection, conspiracy mentality, and political orientation.

Ethical Considerations

The study was approved by the University of Calgary Conjoint Faculty Research Ethics Board (REB20-1178) and was conducted according to the principles expressed in the Declaration of Helsinki.

Measures

Attitudes were measured using three items (“Masking in public is necessary; good; beneficial”) rated on 7-point scales (1=*strongly disagree*, 7=*strongly agree*). Participants indicated their intention to wear a mask in public using a single item (“Over the next months, how often do you intend to wear a face mask when in public?”) also on a 7-point scale (1=*never*, 7=*all the time*). Since the attitude and intention measures were highly correlated, we combined them into a single index of disposition toward masking in public ($\alpha=.96$ preexposure and $\alpha=.97$ postexposure).

To measure perceived objectivity of truth, we asked respondents to consider the question “Should people wear masks in public?” and indicate the extent to which they think there is an objectively true answer to this question [53,54]. They reported their answers on a 7-point scale (1=*definitely no objective truth*, 7=*definitely an objective truth*).

Perceived argument strength was measured using five items adapted from Zhao et al [55]. A sample item is “The arguments of citizen-health are a convincing reason against masking in public” (1=*strongly disagree*, 7=*strongly agree*). The complete scale ($\alpha=.86$ for *citizen-health* and $\alpha=.94$ for *Health_Scientist*) is detailed in [Multimedia Appendix 1](#).

Participants rated the warmth and competence of both protagonists on 11-point bipolar scales adapted from previous research [40,56]. Perceived warmth was assessed using four items ($\alpha=.93$ for *citizen-health* and $\alpha=.96$ for *Health_Scientist*), including *unfriendly/friendly*, *cold/warm*, *irritable/good-natured*, *unsympathetic/sympathetic*. Perceived competence was assessed using six items ($\alpha=.97$ for *citizen-health* and $\alpha=.98$ for *Health_Scientist*), including *uninformed/informed*, *unqualified/qualified*, *unreliable/reliable*, *unbelievable/believable*, and *incompetent/competent*.

Respondents indicated the probability they will be infected with the coronavirus in the next 12 months on a sliding scale (0=0%, 100=100%), and rated how harmful it would be for their health if they were to become infected (1=*not at all*, 5=*extremely*). We computed a perceived COVID-19 risk score by multiplying the probability of infection by the perceived harm and dividing by 100.

Conspiracy mentality was measured using five items ($\alpha=.89$) rated on 7-point scales (1=*strongly disagree*, 7=*strongly agree*) adapted from Bruder et al [57]. A sample item is “Events which superficially seem to lack a connection are often the result of secret activities.”

We assessed cognitive reflection by combining the three items of Frederick’s [58] original Cognitive Reflection Test (CRT; eg, “A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost?”) with the four items of Thomson and Oppenheimer’s [59] nonnumeric CRT (eg, “If you’re running a race and you pass the person in second place, what place are you in?”). Answers were coded 1 for a correct answer and 0 for an incorrect answer. The final cognitive reflection score is the sum of all the correct answers.

The following two items ($r=0.89$) adapted from Schmid and Betsch [40] were used to measure political orientation: (1) If you think about your own political views, where would you classify your views on this scale? (1=*very conservative*, 7=*very liberal*), and (2) If you think about your own political identity, where would you classify your views on this scale? (1=*Republican*, 7=*Democrat*). Scores were reversed so that higher scores indicate political conservatism.

Participants rated their sharing intention on a single item: “If you were to see this post on social media, would you consider sharing it?” (1=*definitely not*, 5=*definitely yes*).

Statistical Analysis

Disposition Toward Masking in Public

Data analysis was performed using the statistical program R, version 4.0 [60], and the level of statistical significance was set at $\alpha=.05$. To answer our main research question, we examined how progressive exposure to false claims and debunking attempts affects people’s attitudes and intentions toward masking in public. Specifically, we tested whether a change in disposition toward masking in public varied from one exposure condition to the next.

Given the structure in our data (each participant provided two sets of ratings), we fit a linear mixed-effects model with disposition toward masking in public as the outcome variable; random intercepts for participants (ID); and fixed effects for exposure condition (contrast-coded using repeated contrasts), time of rating (contrast-coded using treatment contrast), and their interaction. We also added perceived COVID-19 risk, cognitive reflection, political orientation, and conspiracy mentality as mean-centered covariates in the model. The model was estimated using maximum likelihood. We compared this model’s goodness of fit to a second model that was identical but did not include the condition \times time interaction term. The likelihood ratio test indicated that model fit improved significantly when the interaction term was present ($\chi^2_3=33.71$, $P<.001$), thus suggesting a significant interaction.

The P values for the mixed-effects model with interaction were estimated via t tests using the Satterthwaite approximations to degrees of freedom. Effect sizes for the fixed effects are indicated by the standardized regression coefficients (β values) and their 95% CIs

Perceived Objectivity of Truth

Should people wear masks in public? We speculated that during extended debates, the reiteration of false information and rebuke of experts might shake people’s confidence, not only in the veracity of any answer to the question but also in the very existence of an objectively true answer. To test this idea, we performed an analysis of covariance (ANCOVA) on perceived objectivity of truth with exposure condition as the independent variable. Our model controlled for preexposure disposition toward masking in public, perceived COVID-19 risk, cognitive reflection, political orientation, and conspiracy mentality.

Perceived Argument Strength

To test whether exposure to extended debates influences perceptions of the strength of *citizen-health's* arguments, we performed an ANCOVA on perceived argument strength, with condition as the independent variable and controlling for the same set of covariates as indicated above.

Mediation Analysis

We examined whether multiple processes may underly the effect of exposure to misinformation and debunking attempts on the change in people's disposition toward masking in public. Specifically, we tested the idea that correcting the original false claim may improve disposition toward masking, in part, by undermining the strength of the arguments put forth by the misinformation purveyor. Yet, further exposure may generate some doubt about the very existence of an objectively true answer, which, in turn, may weaken people's attitudes and intentions toward masking.

We performed two parallel mediation analyses. The first focused on the difference between the MC and M conditions, whereas the second focused on the difference between the MCR and MC conditions. The mediation models included the change in

disposition toward masking in public as the dependent variable, exposure condition as the independent variable, perceived objectivity as the first mediator, and perceived argument strength as the second mediator. Change in disposition toward masking in public was computed by subtracting participants' initial disposition scores from their postexposure scores. The models were estimated using maximum likelihood with robust standard errors.

Results

Sample Characteristics

A total of 500 participants responded to the survey. Four participants failed the initial attention check and 17 failed the comprehension check. Of those, 6 were in the MC condition, 3 were in the MCR condition, and 8 were in the MCRC condition. After removing responses from participants who failed at least one attention check, we were left with a final sample of 479 participants. The sample's demographic characteristics are shown in [Table 1](#). It is worth noting that including data from participants who failed the attention or comprehension checks did not materially change the size, direction, or statistical significance of the reported effects.

Table 1. Sample characteristics (N=479).

Characteristic	Value
Age (years), mean (SD)	32.1 (12.3)
Gender, n (%)	
Female	257 (53.7)
Male	212 (44.6)
Other	8 (1.7)
Prefer not to answer	2 (0.4)
Education, n (%)	
Less than high school	8 (1.7)
High school graduate	55 (11.5)
Some college but no degree	124 (25.9)
Associate degree	42 (8.8)
Bachelor's degree	158 (33.0)
Master's degree	73 (15.2)
Doctoral degree	3 (0.6)
Professional degree	16 (3.3)
Employment, n (%)	
Employed full time	183 (38.2)
Employed part time	101 (21.1)
Unemployed looking for work	61 (12.7)
Unemployed not looking for work	28 (5.8)
Retired	17 (3.5)
Student	80 (16.7)
Disabled	9 (1.9)

Disposition Toward Masking in Public

Figure 2 shows the estimated marginal means and their 95% CIs. The means and SDs for all variables are presented in Table S1 of Multimedia Appendix 1. Prior to reviewing the Reddit threads, participants in all conditions reported similarly high dispositions toward masking in public. The estimated marginal means did not differ significantly between successive conditions, as shown in Table 2 (ie, *P* values for MC vs M, MCR vs MC, and MCRC vs MCR are all greater than .05). This confirmed that random assignment produced groups with equivalent baselines. The model statistics are summarized in Table 3. Furthermore, exposure to misinformation without any correction resulted in lower disposition toward masking in public (Time 2 in Table 2), while correcting false information improved disposition toward masking. Indeed, participants in the MC condition expressed more positive attitudes and intentions at time 2 than those in the M condition (Time 2: MC vs M in Table 2). Interestingly, the positive effect of correction seemed to vanish when the source of misinformation rebukes the correction

and doubles down on the false claim. Participants in the MCR condition reported significantly lower disposition scores than those in the MC condition (Time 2: MCR vs MC in Table 2). Perhaps even more concerning, a second round of corrections did not appear to effectively counter the impact of the rebuke and doubling down. Indeed, disposition toward masking in public did not differ significantly between participants in the MCRC and MCR conditions.

Although not our primary focus, it is worth noting that all four covariates had significant effects in the expected directions. Disposition toward masking in public was positively related to perceived COVID-19 risk and cognitive reflection (CRT). The latter is consistent with recent research showing CRT to be negatively correlated with the perceived accuracy of fake news, and positively correlated with the ability to discern fake news from real news [61]. Conversely, disposition toward masking in public was negatively related to conspiratorial thinking and political conservatism.

Figure 2. Estimated marginal means and 95% CIs for disposition toward masking in public across conditions and measurement times. M: misinformation-only experimental condition; MC: misinformation+correction experimental condition; MCR: misinformation+correction+rebuke experimental condition; MCRC: misinformation+correction+rebuke+second correction experimental condition.

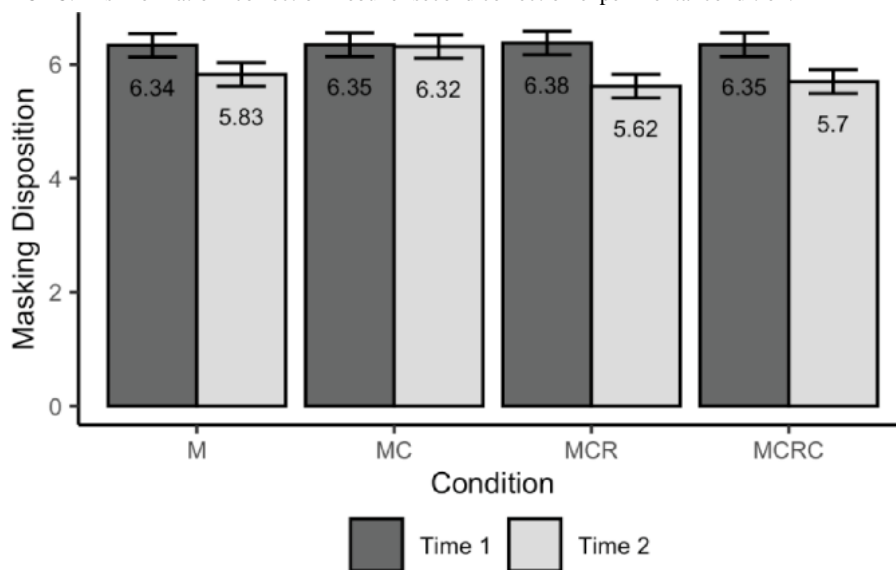


Table 2. Fixed effects for disposition toward masking in public from the mixed-effects regression model.

Predictors	Estimate, <i>b</i> (SE)	<i>t</i> (<i>df</i> =479)	<i>P</i> value	β (95% CI)
Intercept	6.33 (0.05)	121.13	<.001	.18 (.10 to .25)
COVID-19 risk	0.15 (0.05)	3.00	.003	.11 (.04 to .18)
CRT ^a	0.06 (0.02)	2.74	.006	.10 (.03 to .17)
Political orientation	-0.31 (0.03)	-10.48	<.001	-.39 (-.46 to -.32)
Conspiracy mentality	-0.12 (0.04)	-2.99	.003	-.11 (-.18 to -.04)
Time 2	-0.49 (0.05)	-10.38	<.001	-.35 (-.42 to -.29)
MC ^b versus M ^c	0.01 (0.15)	0.07	.94	.01 (-.20 to -.22)
MCR ^d versus MC	0.03 (0.15)	0.22	.83	.02 (-.19 to -.24)
MCRC ^e versus MCR	-0.03 (0.15)	-0.21	.83	-.02 (-.24 to .19)
Time 2: MC versus M	0.48 (0.13)	3.63	<.001	.35 (.16 to .54)
Time 2: MCR versus MC	-0.73 (0.13)	-5.48	<.001	-.53 (-.72 to -.34)
Time 2: MCRC versus MCR	0.11 (0.13)	0.84	.40	.08 (-.11 to .27)

^aCRT: Cognition Reflection Test.

^bMC: misinformation+correction.

^cM: misinformation only.

^dMCR: misinformation+correction+rebuke.

^eMCRC: misinformation+correction+rebuke+second correction.

Table 3. Random effects of the mixed-effects regression model for disposition toward masking in public.

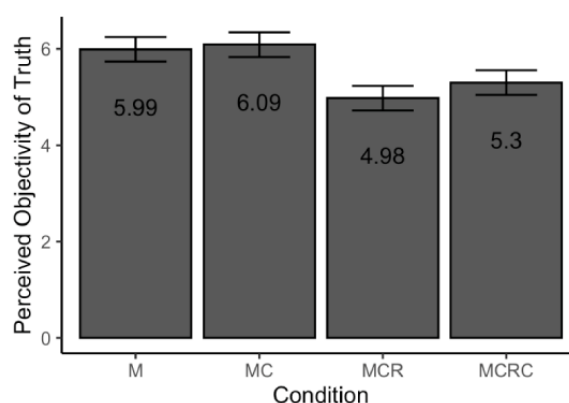
Random effect	Value
σ^2	0.53
τ_{00id}	0.78
Intracorrelation coefficient	0.60
N_{id}	479
Observations	958
Marginal R^2 /Conditional R^2	0.308/0.721

Perceived Objectivity of Truth

The patterns in our data (see the estimated marginal means in Figure 3) lend support to the idea that reiteration of false

information and rebuke of experts might weaken people’s confidence in the very existence of an objectively true answer.

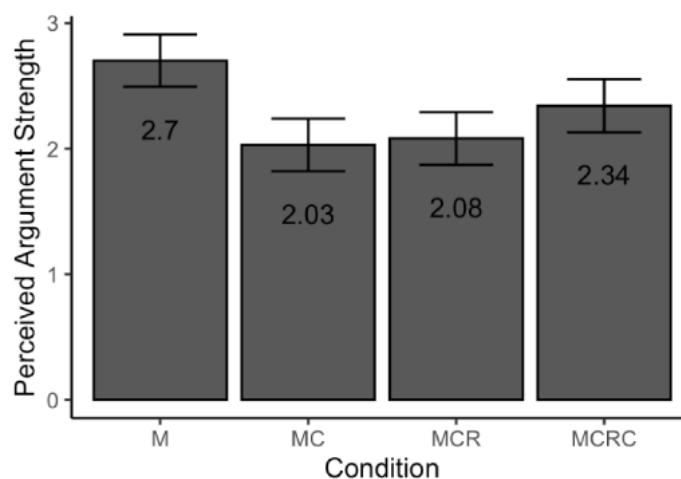
Figure 3. Estimated marginal means and 95% CIs for perceived objectivity of truth across conditions. M: misinformation-only experimental condition; MC: misinformation+correction experimental condition; MCR: misinformation+correction+rebuke experimental condition; MCRC: misinformation+correction+rebuke+second correction experimental condition.



We found a significant effect of condition ($F_{3,470}=16.96, P<.001, \eta_p^2=0.098$). Planned contrasts with Bonferroni correction for multiple tests revealed that people's perception of the objectivity of truth did not change significantly between the MC and M conditions (mean 6.18, SD 1.14 vs mean 5.91, SD 1.34, respectively; $t_{470}=0.53, P>.99, d=0.22, 95\% \text{ CI } -0.03 \text{ to } 0.47$). However, exposure to a second round of misinformation (MCR condition) resulted in appreciably lower perceptions of truth objectivity compared to those for the MC condition (mean 4.96, SD 1.88; $t_{470}=-6.01, P<.001, d=-0.79, 95\% \text{ CI } -1.05 \text{ to } -0.52$). Moreover, witnessing a second correction (MCRC condition) failed to improve perceptions of truth objectivity compared to the MCR condition (mean 5.31, SD 1.81; $t_{470}=1.74, P=.25, d=0.19, 95\% \text{ CI } -0.07 \text{ to } 0.44$). These results suggest that once undermined, perceived objectivity of truth may be difficult to restore.

With respect to the covariates, we found that initial disposition toward masking was positively related to perceived objectivity ($b=0.31, \text{ SE } 0.06; F_{1,470}=25.88, P<.001, \eta_p^2=0.05$), whereas political conservatism ($b=-0.17, \text{ SE } 0.04; F_{1,470}=14.48, P<.001, \eta_p^2=0.03$) and conspiracy mentality ($b=-0.12, \text{ SE } 0.05; F_{1,470}=5.06, P=.02, \eta_p^2=0.01$) were negatively related to perceived objectivity. The effects of cognitive reflection ($b=0.03, \text{ SE } 0.03; F_{1,470}=0.92, P=.34, \eta_p^2=0.002$) and perceived COVID-19 risk ($b=0.10, \text{ SE } 0.07; F_{1,470}=1.96, P=.16, \eta_p^2=0.004$) were not significant.

Figure 4. Estimated marginal means and 95% CIs for perceived argument strength across exposure conditions. M: misinformation-only experimental condition; MC: misinformation+correction experimental condition; MCR: misinformation+correction+rebuke experimental condition; MCRC: misinformation+correction+rebuke+second correction experimental condition.



As for the covariates, political conservatism ($b=0.14, \text{ SE } 0.04; F_{1,470}=14.03, P<.001, \eta_p^2=0.029$) and conspiracy mentality ($b=0.16, \text{ SE } 0.04; F_{1,470}=12.19, P<.001, \eta_p^2=0.025$) were positively related to perceived argument strength, whereas initial disposition to wearing masks ($b=-0.39, \text{ SE } 0.05; F_{1,470}=62.96, P<.001, \eta_p^2=0.118$) and cognitive reflection ($b=-0.05, \text{ SE } 0.03; F_{1,470}=4.13, P=.04, \eta_p^2=0.009$) were negatively related to perceived argument strength. Perceived COVID-19 risk ($b=0.08,$

Perceived Argument Strength

The previous analysis suggested that reduction in the perceived objectivity of truth only occurs following exposure to the second round of misinformation. Thus, perceived objectivity of truth cannot account for the observed changes in attitude and intention across the entire range of exposure conditions. In particular, it cannot account for the improvement in disposition toward masking following correction of the initial false claim. Therefore, we next examined whether perceptions of the strength of *citizen-health's* arguments may provide an alternative account.

Exposure condition had a significant impact on perceived argument strength ($F_{3,470}=8.21, P<.001, \eta_p^2=0.05$). Planned contrasts with Bonferroni correction for multiple tests revealed that *citizen-health's* arguments were perceived to be weaker in the MC condition than in the M condition (mean 1.95, SD 1.26 vs mean 2.75, SD 1.59, respectively; $t_{470}=-4.43, P<.001, d=-0.55, 95\% \text{ CI } -0.81 \text{ to } -0.30$). However, perceived argument strength remained low in the MCR condition and showed no difference from that in the MC condition (mean 2.12, SD 1.31; $t_{470}=0.34, P>.99, d=0.13, 95\% \text{ CI } -0.12 \text{ to } 0.39$). It also did not differ significantly between the MCRC condition and the MCR condition (mean 2.34, SD 1.30; $t_{470}=1.72, P=.26, d=0.17, 95\% \text{ CI } -0.09 \text{ to } 0.42$). The estimated marginal means are shown in Figure 4.

$b=0.08, \text{ SE } 0.06; F_{1,470}=1.92, P=.17, \eta_p^2=0.004$) was unrelated to perceived argument strength.

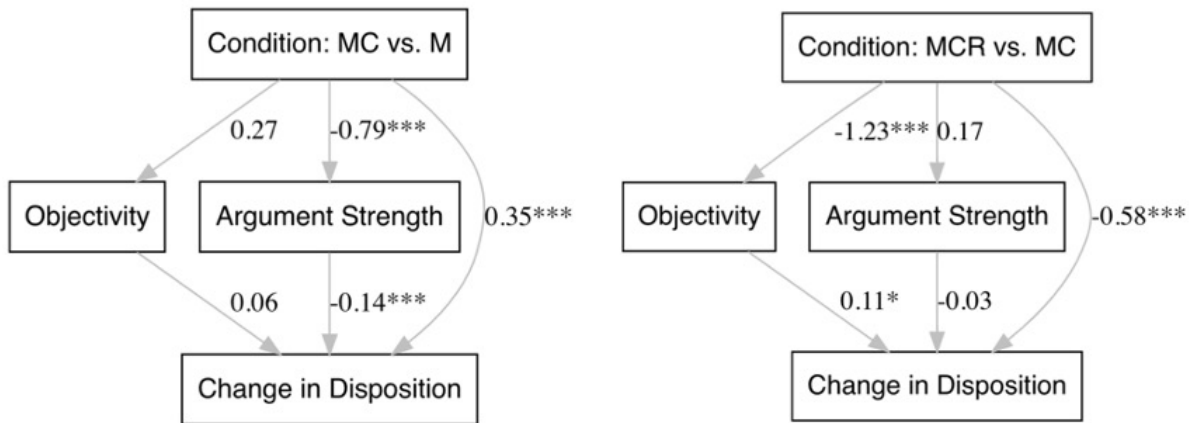
Mediation Analysis

When comparing the MC and M conditions, the effect of increased exposure on change in disposition toward masking in public was partially mediated by perceived argument strength (Figure 5). Indeed, both the indirect effect Condition through Argument Strength through Change in Disposition ($b=0.12, \text{ SE } 0.04, z=2.96; P=.01, 95\% \text{ CI } 0.04-0.19$) and the direct effect

Condition through Change in Disposition ($b=0.35$, SE 0.09, $z=3.90$; $P<.001$, 95% CI 0.17-0.52) were significant. However, the second indirect effect Condition through Objectivity through Change in Disposition was not significant ($b=0.02$, SE 0.02, $z=0.93$; $P=.35$, 95% CI -0.02 to 0.05). Moreover, a formal test

of the difference between the indirect effects confirmed that the indirect effect through Argument Strength was significantly larger than the indirect effect through Perceived Objectivity ($b=0.10$, SE 0.04, $z=2.84$; $P=.004$, 95% CI 0.03-0.17).

Figure 5. Objectivity and argument strength mediate the effect of exposure on change in disposition. Values are unstandardized coefficients (b values). M: misinformation-only experimental condition; MC: misinformation+correction experimental condition; MCR: misinformation+correction+rebutal experimental condition. * $P<.05$, *** $P<.001$.



Conversely, when considering the MCR and MC conditions, perceived objectivity partially mediated the effect of exposure on change in disposition. The indirect effect Condition through Objectivity through Change in Disposition was statistically significant ($b=-0.14$, SE 0.07, $z=-2.09$; $P=.04$, 95% CI -0.27 to -0.01), as was the direct effect Condition through Change in Disposition ($b=-0.58$, SE 0.13, $z=-4.59$; $P<.001$, 95% CI -0.83 to -0.33). Moreover, the indirect effect Condition through Argument Strength through Change in Disposition was not significant ($b=-0.004$, SE 0.01, $z=-0.47$; $P=.64$, 95% CI -0.02 to 0.01), and the indirect effect through Perceived Objectivity was larger than the indirect effect through Argument Strength ($b=0.14$, SE 0.07, $z=2.00$; $P=.05$, 95% CI 0.003-0.27).

it is likely that there are additional mediators that could contribute to understanding the dynamic effects of extended exposure to misinformation on people’s attitudes and intentions toward behaviors favored by science.

These results suggest that no single process can fully account for the observed patterns in the data. Although a decrease in perceived argument strength partially explains why seeing a correction following initial exposure to misinformation improves attitudes and intentions toward masking, this path accounted for only 24.0% of the total effect of seeing a correction on change in attitudes and intentions. Moreover, a decrease in the perceived objectivity of truth partially explains why exposure to a second round of misinformation that includes rebuttal of the correction and doubling down on the original false claim negatively impacts attitudes and intentions toward masking. However, this path accounted for only 19.4% of the total effect of further exposure to misinformation on change in disposition toward masking.

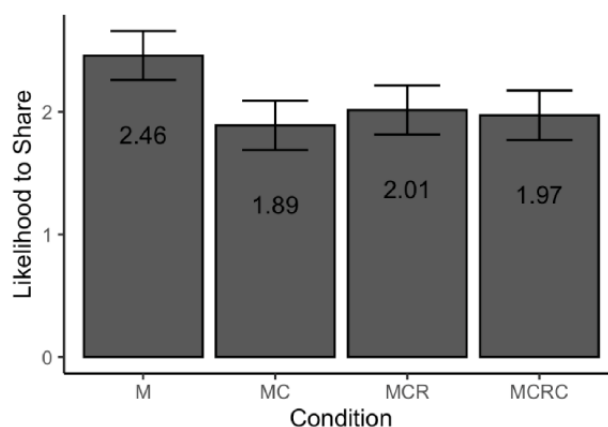
Additional Analyses

We tested for other potential mediators, but found that increased exposure did not influence participants’ perceptions of *citizen-health’s* warmth ($F_{3,470}=1.73$, $P=.16$, $\eta_p^2=0.01$), *citizen-health’s* competence ($F_{3,470}=0.92$, $P=.43$, $\eta_p^2=0.006$), *Health_Scientist’s* warmth ($F_{2,349}=0.63$, $P=.53$, $\eta_p^2=0.004$), *Health_Scientist’s* competence ($F_{2,349}=0.37$, $P=.69$, $\eta_p^2=0.002$), or strength of *Health_Scientist’s* arguments ($F_{2,349}=1.39$, $P=.25$, $\eta_p^2=0.008$).

Finally, we analyzed participants’ intention to share misinformation on social media (see Figure 6). Exposure had a significant effect on sharing intention ($F_{3,470}=6.24$, $P<.001$, $\eta_p^2=0.038$). Planned contrasts with Bonferroni correction for multiple comparisons showed a decrease in the intention to share the original posting after seeing a correction (MC mean 1.88, SD 1.15 vs M mean 2.48, SD 1.40; $t_{470}=-3.93$, $P<.001$, $d=-0.47$, 95% CI -0.72 to -0.21). Furthermore, there was no significant difference in sharing intention between the MCR and MC conditions (mean 2.03, SD 0.99; $t_{470}=0.86$, $P>.99$, $d=0.14$, 95% CI -0.11 to 0.39), or between the MCRC and MCR conditions (mean 1.96, SD 1.09; $t_{470}=-0.29$, $P>.99$, $d=-0.07$, 95% CI -0.32 to 0.19).

Given the significant direct effect of exposure on change in disposition and the modest proportion mediated in both cases,

Figure 6. Marginal means and 95% CIs of intention to share across exposure conditions. M: misinformation-only experimental condition; MC: misinformation+correction experimental condition; MCR: misinformation+correction+rebuttal experimental condition; MCRC: misinformation+correction+rebuttal+second correction experimental condition.



Examining the covariates, we found that initial disposition to wearing masks ($b=-0.23$, SE 0.05; $F_{1,470}=23.76$, $P<.001$, $\eta_p^2=0.046$) and cognitive reflection ($b=-0.06$, SE 0.02; $F_{1,470}=6.32$, $P=.01$, $\eta_p^2=0.013$) were negatively associated with intention to share. Neither perceived COVID-19 risk ($b=0.06$, SE 0.05; $F_{1,470}=1.33$, $P=.25$, $\eta_p^2=0.003$) nor political orientation ($b=0.04$, SE 0.03; $F_{1,470}=1.55$, $P=.21$, $\eta_p^2=0.003$) or conspiracy mentality ($b=0.01$, SE 0.04; $F_{1,470}=0.03$, $P=.86$, $\eta_p^2=0.000$) was significantly related to sharing intentions.

These results indicate that people are more likely to share misinformation when its content is consistent with their existing beliefs about the issue. Importantly, they also suggest that correcting a false claim can reduce the extent of its spread on social media, and this effect seems resistant to further exposure to the same misinformation.

Discussion

Principal Findings

Past research from across a variety of domains has shown that debunking misinformation works. That is, well-crafted corrections delivered by trusted sources can positively impact beliefs, attitudes, and intentions toward behavior favored by science [26-36]. With the explosion of false and misleading health claims on social media, especially since the start of the COVID-19 pandemic [62,63], scientists, experts, and health care professionals have been called upon to increase their presence on social media and help combat this “infodemic” [20-27]. A recent study found that US physicians and nurses are generally willing to take on the task even if it comes with important challenges [64]. However, little is known about the impact of extended social media debates on observers’ attitudes and intentions toward the debated issue. In this study, we tested such an impact in the context of a debate about the safety and effectiveness of wearing face masks in public during the early days of the pandemic.

We found that exposure to misinformation has a negative impact on attitudes and intentions toward masking. This result is consistent with prior research finding that exposure to

misinformation negatively impacts attitudes and intentions toward behaviors favored by science [40].

Also in line with prior work [28-31], we found that initial debunking of a false claim generally improves attitudes and intentions toward masking. This effect is partially explained by a decrease in the perceived strength of the argument underlying the false claim. However, this improvement is washed out by further exposure to false claims and debunking attempts. The latter result is partially explained by a decrease in the perceived objectivity of truth. That is, extended exposure to false claims and debunking attempts appears to weaken the belief that there is a nonsubjective, correct answer to how people ought to behave in this situation, which in turn leads to less positive reactions toward masking as the prescribed behavior. Interestingly, exposure to contradictory information affects perceived truth objectivity in a nonlinear fashion. For instance, exposure to a false claim and its initial debunking does not weaken the belief that there is an objectively true answer. It appears that the level of exposure to contradictory information needs to reach a certain threshold before it affects perceived truth objectivity.

Finally, we found that people are more likely to share misinformation when its content is consistent with their existing beliefs. However, correcting misinformation reduces its likelihood of being shared on social media, and this effect persists even after multiple exposures. These results, while highlighting the value of debunking in combating the spread of misinformation on social media, suggest that, unlike attitudes, sharing intentions may be insensitive to extended exposure to a back and forth between misinformation and correction. This pattern may reflect a floor effect, in that people had expressed very low intentions to share corrected misinformation (mean of 1.89 on a 7-point scale). Exposure to further debate resulted in more negative attitudes toward masking, but did not impact sharing intentions because sharing intentions were already extremely low.

Comparison With Prior Work

Our findings have important implications for research on debunking misinformation. Extant literature has noted that even though corrections generally reduce people’s beliefs in false information, the misinformation often continues to influence

their thinking, a phenomenon known as the continued influence effect [65]. Once people process information that appears vaguely credible to them, it becomes difficult to retract it. A popular explanation of the continued influence effect assumes that people build mental models of the world and want their models to be complete. They are willing to accept false information if it allows them to build complete models. When that false information is later debunked, it creates a gap in their understanding of the world. Since people dislike gaps in their understanding and prefer their mental models to be complete, they continue to rely on information they know is false. [65,66].

Another explanation of the continued influence effect argues that attempts to correct misinformation often end up reinforcing it through repetition [67,68]. From this perspective, repeating misinformation when attempting to correct it makes it feel more familiar and fluent. By inadvertently increasing the ease with which misinformation is processed, correction attempts also increase the likelihood of it being accepted as true. Our study suggests yet another possible explanation of the stickiness of misinformation. In some situations, witnessing a heated debate with arguments for and against a controversial issue could undermine people's confidence in the existence of an objectively true answer, which may weaken their commitment to either side of the debate.

Limitations and Future Research

We set out to study the impact of exposure to extended debates on social media. However, our study was limited to a single platform (Reddit), and the debate was restricted to four exchanges between only two protagonists. This limits the generalizability of our findings. Interaction norms likely differ across social media platforms, which may impact how users interpret the conversation. Future research could attempt to replicate our findings using different social media platforms (eg, Facebook and Twitter), and examine the consequences of extending the debate to include more than two protagonists and more than four exchanges. Relatedly, although extended debates such as those described in this research are familiar to regular users of social media, we do not know how often they happen. Future research would benefit from quantifying the frequency of their occurrence and how it may vary across platforms.

A possible limitation of our experimental design is that the rebuke message included an image, whereas all other messages were strictly text-based. Thus, it is impossible to disentangle the impact of exposure to the rebuke message from the presence of an image. However, pinpointing which message element (image vs text) accounts for the effect of exposure was not a goal of this study. Instead, we sought to pit strong refutation against persuasive misinformation. We chose to include a graphic image in the rebuke message precisely because of its persuasive power.

Although the debunking messages used in this study were developed using recommended best practices, it is possible that different debunking techniques would have resulted in different outcomes. Future research would greatly benefit from manipulating features of the debunking argument as well as the source of debunking. For example, previous research has shown that debunking messages from anonymous social media users are more effective when they include a link to a trusted source such as the CDC [26]. Future research could test whether the positive impact of providing links to credible sources persists in the context of extended social media debates. Other features of debunking messages previously found to sometimes reduce misperceptions include the use of humor [69] and infographics [70]. However, whether such debunking techniques can be effective in the context of extended debates remains an open question.

Preemptively refuting misinformation or even just warning people that they might be misinformed has been shown to decrease later reliance on misinformation. Future research could test the effectiveness of such prebunking in the context of extended debates, where—to extend the biomedical analogy—the viral load is higher.

Conclusions

In sum, dynamic conversations present a heretofore underappreciated challenge faced by health professionals and science advocates attempting to debunk misinformation on social media. Engaging in extended debates with science deniers and other purveyors of bunk appears necessary, but more research is needed to address the unintended consequences of such engagement.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Complete scales for measures and descriptive statistics.
[DOCX File, 27 KB - [jmir_v24i3e34831_app1.docx](#)]

References

1. Fox S. The social life of health information. Pew Research Center. 2014 Jan 15. URL: <https://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/> [accessed 2021-09-21]
2. Mitchell A, Liedke J. About four-in-ten Americans say social media is an important way of following COVID-19 vaccine news. Pew Research Center. 2021 Aug 24. URL: <https://www.pewresearch.org/fact-tank/2021/08/24/about-four-in-ten-americans-say-social-media-is-an-important-way-of-following-covid-19-vaccine-news/> [accessed 2021-09-30]

3. Sylvia Chou W, Gaysynsky A, Cappella JN. Where we go from here: health misinformation on social media. *Am J Public Health* 2020 Oct;110(S3):S273-S275. [doi: [10.2105/ajph.2020.305905](https://doi.org/10.2105/ajph.2020.305905)]
4. Southwell BG, Thorson EA, Sheble L. The persistence and peril of misinformation. *Am Sci* 2017;105(6):368-371. [doi: [10.1511/2017.105.6.372](https://doi.org/10.1511/2017.105.6.372)]
5. 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](https://doi.org/10.1126/science.aap9559)] [Medline: [29590045](https://pubmed.ncbi.nlm.nih.gov/29590045/)]
6. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. *J Med Internet Res* 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: [10.2196/17187](https://doi.org/10.2196/17187)] [Medline: [33470931](https://pubmed.ncbi.nlm.nih.gov/33470931/)]
7. 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](https://doi.org/10.1016/j.socscimed.2019.112552)] [Medline: [31561111](https://pubmed.ncbi.nlm.nih.gov/31561111/)]
8. Cummings M. Misinformation on Reddit has become unmanageable, Alberta moderators say. *CBC News*. 2021 Sep 21. URL: <https://www.cbc.ca/news/canada/edmonton/misinformation-alberta-reddit-unmanageable-moderators-1.6179120> [accessed 2021-09-21]
9. 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](https://doi.org/10.2105/AJPH.2018.304567)] [Medline: [30138075](https://pubmed.ncbi.nlm.nih.gov/30138075/)]
10. Tan ASL, Lee C, Chae J. Exposure to health (mis)information: lagged effects on young adults' health behaviors and potential pathways. *J Commun* 2015 Jul 06;65(4):674-698. [doi: [10.1111/jcom.12163](https://doi.org/10.1111/jcom.12163)]
11. Ratzan SC. Setting the record straight: vaccines, autism, and the Lancet. *J Health Commun* 2010 Apr 27;15(3):237-239. [doi: [10.1080/10810731003780714](https://doi.org/10.1080/10810731003780714)] [Medline: [20432106](https://pubmed.ncbi.nlm.nih.gov/20432106/)]
12. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: cross-sectional online study. *J Med Internet Res* 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: [10.2196/22205](https://doi.org/10.2196/22205)] [Medline: [33048825](https://pubmed.ncbi.nlm.nih.gov/33048825/)]
13. Luk TT, Zhao S, Weng X, Wong JY, Wu YS, Ho SY, et al. Exposure to health misinformation about COVID-19 and increased tobacco and alcohol use: a population-based survey in Hong Kong. *Tob Control* 2021 Nov;30(6):696-699. [doi: [10.1136/tobaccocontrol-2020-055960](https://doi.org/10.1136/tobaccocontrol-2020-055960)] [Medline: [32855353](https://pubmed.ncbi.nlm.nih.gov/32855353/)]
14. Wilson SL, Wysonge C. Social media and vaccine hesitancy. *BMJ Glob Health* 2020 Oct;5(10):e004206 [FREE Full text] [doi: [10.1136/bmjgh-2020-004206](https://doi.org/10.1136/bmjgh-2020-004206)] [Medline: [33097547](https://pubmed.ncbi.nlm.nih.gov/33097547/)]
15. Pennycook G, Rand DG. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc Natl Acad Sci U S A* 2019 Feb 12;116(7):2521-2526 [FREE Full text] [doi: [10.1073/pnas.1806781116](https://doi.org/10.1073/pnas.1806781116)] [Medline: [30692252](https://pubmed.ncbi.nlm.nih.gov/30692252/)]
16. Lewandowsky S, Cook J, Ecker U, Albarracín D, Amazeen M, Kendeou P, et al. The Debunking Handbook 2020. *Skeptical Science*. 2020 Oct 14. URL: <https://sks.to/db2020> [accessed 2021-09-30]
17. Kozyreva A, Lewandowsky S, Hertwig R. Citizens versus the internet: confronting digital challenges with cognitive tools. *Psychol Sci Public Interest* 2020 Dec;21(3):103-156 [FREE Full text] [doi: [10.1177/1529100620946707](https://doi.org/10.1177/1529100620946707)] [Medline: [33325331](https://pubmed.ncbi.nlm.nih.gov/33325331/)]
18. Lazer DMJ, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, et al. The science of fake news. *Science* 2018 Mar 09;359(6380):1094-1096. [doi: [10.1126/science.aao2998](https://doi.org/10.1126/science.aao2998)] [Medline: [29590025](https://pubmed.ncbi.nlm.nih.gov/29590025/)]
19. Pennycook G, Epstein Z, Mosleh M, Arechar AA, Eckles D, Rand DG. Shifting attention to accuracy can reduce misinformation online. *Nature* 2021 Apr 17;592(7855):590-595. [doi: [10.1038/s41586-021-03344-2](https://doi.org/10.1038/s41586-021-03344-2)] [Medline: [33731933](https://pubmed.ncbi.nlm.nih.gov/33731933/)]
20. Gunter J. Medical misinformation and the internet: a call to arms. *Lancet* 2019 Jun 08;393(10188):2294-2295. [doi: [10.1016/S0140-6736\(19\)31206-1](https://doi.org/10.1016/S0140-6736(19)31206-1)] [Medline: [31180024](https://pubmed.ncbi.nlm.nih.gov/31180024/)]
21. Armstrong PW, Naylor CD. Counteracting health misinformation: a role for medical journals? *JAMA* 2019 May 21;321(19):1863-1864. [doi: [10.1001/jama.2019.5168](https://doi.org/10.1001/jama.2019.5168)] [Medline: [31009036](https://pubmed.ncbi.nlm.nih.gov/31009036/)]
22. Pagoto S, Waring ME, Xu R. A call for a public health agenda for social media research. *J Med Internet Res* 2019 Dec 19;21(12):e16661 [FREE Full text] [doi: [10.2196/16661](https://doi.org/10.2196/16661)] [Medline: [31855185](https://pubmed.ncbi.nlm.nih.gov/31855185/)]
23. O'Connor C, Murphy M. Going viral: doctors must tackle fake news in the covid-19 pandemic. *BMJ* 2020 Apr 24;369:m1587. [doi: [10.1136/bmj.m1587](https://doi.org/10.1136/bmj.m1587)] [Medline: [32332066](https://pubmed.ncbi.nlm.nih.gov/32332066/)]
24. Rubin R. Getting social: physicians can counteract misinformation with an online presence. *JAMA* 2019 Aug 20;322(7):598-600. [doi: [10.1001/jama.2019.10779](https://doi.org/10.1001/jama.2019.10779)] [Medline: [31365030](https://pubmed.ncbi.nlm.nih.gov/31365030/)]
25. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. *Annu Rev Public Health* 2020 Apr 02;41(1):433-451. [doi: [10.1146/annurev-publhealth-040119-094127](https://doi.org/10.1146/annurev-publhealth-040119-094127)] [Medline: [31874069](https://pubmed.ncbi.nlm.nih.gov/31874069/)]
26. Vraga EK, Bode L. I do not believe you: how providing a source corrects health misperceptions across social media platforms. *Inf Commun Soc* 2017 Apr 19;21(10):1337-1353. [doi: [10.1080/1369118x.2017.1313883](https://doi.org/10.1080/1369118x.2017.1313883)]
27. Vraga EK, Bode L. Using expert sources to correct health misinformation in social media. *Sci Commun* 2017 Sep 14;39(5):621-645. [doi: [10.1177/1075547017731776](https://doi.org/10.1177/1075547017731776)]

28. Chan MS, Jones CR, Hall Jamieson K, Albarracín D. Debunking: a meta-analysis of the psychological efficacy of messages countering misinformation. *Psychol Sci* 2017 Nov;28(11):1531-1546 [FREE Full text] [doi: [10.1177/0956797617714579](https://doi.org/10.1177/0956797617714579)] [Medline: [28895452](https://pubmed.ncbi.nlm.nih.gov/28895452/)]
29. Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J. Misinformation and its correction: continued influence and successful debiasing. *Psychol Sci Public Interest* 2012 Dec;13(3):106-131. [doi: [10.1177/1529100612451018](https://doi.org/10.1177/1529100612451018)] [Medline: [26173286](https://pubmed.ncbi.nlm.nih.gov/26173286/)]
30. Walter N, Tukachinsky R. A meta-analytic examination of the continued influence of misinformation in the face of correction: how powerful is it, why does it happen, and how to stop it? *Commun Res* 2019 Jun 22;47(2):155-177. [doi: [10.1177/0093650219854600](https://doi.org/10.1177/0093650219854600)]
31. Vraga EK, Bode L. Correction as a solution for health misinformation on social media. *Am J Public Health* 2020 Oct;110(S3):S278-S280. [doi: [10.2105/AJPH.2020.305916](https://doi.org/10.2105/AJPH.2020.305916)] [Medline: [33001724](https://pubmed.ncbi.nlm.nih.gov/33001724/)]
32. Bode L, Vraga EK. See something, say something: correction of global health misinformation on social media. *Health Commun* 2018 Sep;33(9):1131-1140. [doi: [10.1080/10410236.2017.1331312](https://doi.org/10.1080/10410236.2017.1331312)] [Medline: [28622038](https://pubmed.ncbi.nlm.nih.gov/28622038/)]
33. Bode L, Vraga EK. In related news, that was wrong: the correction of misinformation through related stories functionality in social media. *J Commun* 2015 Jun 23;65(4):619-638. [doi: [10.1111/jcom.12166](https://doi.org/10.1111/jcom.12166)]
34. van der Meer TGLA, Jin Y. Seeking formula for misinformation treatment in public health crises: the effects of corrective information type and source. *Health Commun* 2020 May 14;35(5):560-575. [doi: [10.1080/10410236.2019.1573295](https://doi.org/10.1080/10410236.2019.1573295)] [Medline: [30761917](https://pubmed.ncbi.nlm.nih.gov/30761917/)]
35. Gesser-Edelsburg A, Diamant A, Hijazi R, Mesch GS. Correcting misinformation by health organizations during measles outbreaks: a controlled experiment. *PLoS One* 2018 Dec 19;13(12):e0209505 [FREE Full text] [doi: [10.1371/journal.pone.0209505](https://doi.org/10.1371/journal.pone.0209505)] [Medline: [30566485](https://pubmed.ncbi.nlm.nih.gov/30566485/)]
36. Vraga EK, Bode L, Tully M. The effects of a news literacy video and real-time corrections to video misinformation related to sunscreen and skin cancer. *Health Commun* 2021 Apr 12:1-9. [doi: [10.1080/10410236.2021.1910165](https://doi.org/10.1080/10410236.2021.1910165)] [Medline: [33840310](https://pubmed.ncbi.nlm.nih.gov/33840310/)]
37. Fischer CB, Adrien N, Silguero JJ, Hopper JJ, Chowdhury AI, Werler MM. Mask adherence and rate of COVID-19 across the United States. *PLoS One* 2021 Apr 14;16(4):e0249891 [FREE Full text] [doi: [10.1371/journal.pone.0249891](https://doi.org/10.1371/journal.pone.0249891)] [Medline: [33852626](https://pubmed.ncbi.nlm.nih.gov/33852626/)]
38. The CoronaVirusFacts/DatosCoronaVirus Alliance Database. Poynter. URL: <https://www.poynter.org/ifcn-covid-19-misinformation/> [accessed 2021-09-30]
39. Paynter J, Luskin-Saxby S, Keen D, Fordyce K, Frost G, Imms C, et al. Evaluation of a template for countering misinformation-Real-world Autism treatment myth debunking. *PLoS One* 2019;14(1):e0210746 [FREE Full text] [doi: [10.1371/journal.pone.0210746](https://doi.org/10.1371/journal.pone.0210746)] [Medline: [30699155](https://pubmed.ncbi.nlm.nih.gov/30699155/)]
40. Schmid P, Betsch C. Effective strategies for rebutting science denialism in public discussions. *Nat Hum Behav* 2019 Sep 24;3(9):931-939. [doi: [10.1038/s41562-019-0632-4](https://doi.org/10.1038/s41562-019-0632-4)] [Medline: [31235861](https://pubmed.ncbi.nlm.nih.gov/31235861/)]
41. Lewandowsky S, van der Linden S. Countering misinformation and fake news through inoculation and prebunking. *Eur Rev Soc Psychol* 2021 Feb 22;32(2):348-384. [doi: [10.1080/10463283.2021.1876983](https://doi.org/10.1080/10463283.2021.1876983)]
42. Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman ALJ, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. *R Soc Open Sci* 2020 Oct 14;7(10):201199 [FREE Full text] [doi: [10.1098/rsos.201199](https://doi.org/10.1098/rsos.201199)] [Medline: [33204475](https://pubmed.ncbi.nlm.nih.gov/33204475/)]
43. Cook J, Lewandowsky S, Ecker UKH. Neutralizing misinformation through inoculation: exposing misleading argumentation techniques reduces their influence. *PLoS One* 2017;12(5):e0175799 [FREE Full text] [doi: [10.1371/journal.pone.0175799](https://doi.org/10.1371/journal.pone.0175799)] [Medline: [28475576](https://pubmed.ncbi.nlm.nih.gov/28475576/)]
44. Lewandowsky S, Ecker UK, Cook J. Beyond misinformation: understanding and coping with the “post-truth” era. *J Appl Res Mem Cogn* 2017 Dec;6(4):353-369. [doi: [10.1016/j.jarmac.2017.07.008](https://doi.org/10.1016/j.jarmac.2017.07.008)]
45. Clapper JR. Facts and fears: hard truths from a life in intelligence. New York: Viking Penguin; 2018.
46. Nagler RH, Yzer MC, Rothman AJ. Effects of media exposure to conflicting information about mammography: results from a population-based survey experiment. *Ann Behav Med* 2019 Aug 29;53(10):896-908 [FREE Full text] [doi: [10.1093/abm/kay098](https://doi.org/10.1093/abm/kay098)] [Medline: [30596830](https://pubmed.ncbi.nlm.nih.gov/30596830/)]
47. Nagler R, Vogel R, Gollust S, Yzer M, Rothman A. Effects of prior exposure to conflicting health information on responses to subsequent unrelated health messages: results from a population-based longitudinal experiment. *Ann Behav Med* 2021 Aug 16:kaab069. [doi: [10.1093/abm/kaab069](https://doi.org/10.1093/abm/kaab069)] [Medline: [34398961](https://pubmed.ncbi.nlm.nih.gov/34398961/)]
48. Lee C, Nagler RH, Wang N. Source-specific exposure to contradictory nutrition information: documenting prevalence and effects on adverse cognitive and behavioral outcomes. *Health Commun* 2018 Apr 02;33(4):453-461 [FREE Full text] [doi: [10.1080/10410236.2016.1278495](https://doi.org/10.1080/10410236.2016.1278495)] [Medline: [28151010](https://pubmed.ncbi.nlm.nih.gov/28151010/)]
49. Chang C. Motivated processing. *Sci Commun* 2015 Jul 30;37(5):602-634. [doi: [10.1177/1075547015597914](https://doi.org/10.1177/1075547015597914)]
50. Nan X, Daily K. Biased assimilation and need for closure: examining the effects of mixed blogs on vaccine-related beliefs. *J Health Commun* 2015 Apr 09;20(4):462-471. [doi: [10.1080/10810730.2014.989343](https://doi.org/10.1080/10810730.2014.989343)] [Medline: [25751250](https://pubmed.ncbi.nlm.nih.gov/25751250/)]
51. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. *J Exp Soc Psychol* 2017 May;70:153-163. [doi: [10.1016/j.jesp.2017.01.006](https://doi.org/10.1016/j.jesp.2017.01.006)]

52. Palan S, Schitter C. Prolific.ac—A subject pool for online experiments. *J Behav Exp Finance* 2018 Mar;17:22-27. [doi: [10.1016/j.jbef.2017.12.004](https://doi.org/10.1016/j.jbef.2017.12.004)]
53. Fisher M, Knobe J, Strickland B, Keil FC. The influence of social interaction on intuitions of objectivity and subjectivity. *Cogn Sci* 2017 May 01;41(4):1119-1134. [doi: [10.1111/cogs.12380](https://doi.org/10.1111/cogs.12380)] [Medline: [27246257](https://pubmed.ncbi.nlm.nih.gov/27246257/)]
54. Sarkissian H, Park J, Tien D, Wright J, Knobe J. Folk moral relativism. *Mind Lang* 2011;26:482-505. [doi: [10.1111/j.1468-0017.2011.01428.x](https://doi.org/10.1111/j.1468-0017.2011.01428.x)]
55. Zhao X, Strasser A, Cappella JN, Lerman C, Fishbein M. A measure of perceived argument strength: reliability and validity. *Commun Methods Meas* 2011 Mar 04;5(1):48-75 [FREE Full text] [doi: [10.1080/19312458.2010.547822](https://doi.org/10.1080/19312458.2010.547822)] [Medline: [25568663](https://pubmed.ncbi.nlm.nih.gov/25568663/)]
56. McCroskey JC, Jenson TA. Image of mass media news sources. *J Broadcasting* 1975 Mar;19(2):169-180. [doi: [10.1080/08838157509363777](https://doi.org/10.1080/08838157509363777)]
57. Bruder M, Haffke P, Neave N, Nouripanah N, Imhoff R. Measuring individual differences in generic beliefs in conspiracy theories across cultures: conspiracy mentality questionnaire. *Front Psychol* 2013;4:225. [doi: [10.3389/fpsyg.2013.00225](https://doi.org/10.3389/fpsyg.2013.00225)] [Medline: [23641227](https://pubmed.ncbi.nlm.nih.gov/23641227/)]
58. Frederick S. Cognitive reflection and decision making. *J Econ Persp* 2005 Nov 01;19(4):25-42. [doi: [10.1257/089533005775196732](https://doi.org/10.1257/089533005775196732)]
59. Thomson K, Oppenheimer D. Investigating an alternate form of the cognitive reflection test. *Judgm Decis Mak* 2016;11(1):99-113 [FREE Full text]
60. The R Project for Statistical Computing Version 4. URL: <https://www.R-project.org/> [accessed 2022-02-19]
61. Pennycook G, Rand DG. Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 2019 Jul;188:39-50. [doi: [10.1016/j.cognition.2018.06.011](https://doi.org/10.1016/j.cognition.2018.06.011)] [Medline: [29935897](https://pubmed.ncbi.nlm.nih.gov/29935897/)]
62. Al-Zaman MS. Prevalence and source analysis of COVID-19 misinformation in 138 countries. *IFLA J* 2021 Aug 27:034003522110411. [doi: [10.1177/03400352211041135](https://doi.org/10.1177/03400352211041135)]
63. Cinelli M, Quattrocioni W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. *Sci Rep* 2020 Oct 06;10(1):16598. [doi: [10.1038/s41598-020-73510-5](https://doi.org/10.1038/s41598-020-73510-5)] [Medline: [33024152](https://pubmed.ncbi.nlm.nih.gov/33024152/)]
64. Bautista JR, Zhang Y, Gwizdzka J. US physicians' and nurses' motivations, barriers, and recommendations for correcting health misinformation on social media: qualitative interview study. *JMIR Public Health Surveill* 2021 Sep 01;7(9):e27715 [FREE Full text] [doi: [10.2196/27715](https://doi.org/10.2196/27715)] [Medline: [34468331](https://pubmed.ncbi.nlm.nih.gov/34468331/)]
65. Johnson HM, Seifert CM. Sources of the continued influence effect: when misinformation in memory affects later inferences. *J Exp Psychol Learn Mem Cogn* 1994 Nov;20(6):1420-1436. [doi: [10.1037/0278-7393.20.6.1420](https://doi.org/10.1037/0278-7393.20.6.1420)]
66. Ecker UKH, Lewandowsky S, Tang DTW. Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Mem Cogn* 2010 Dec;38(8):1087-1100. [doi: [10.3758/mc.38.8.1087](https://doi.org/10.3758/mc.38.8.1087)]
67. Skurnik I, Yoon C, Park D, Schwarz N. How warnings about false claims become recommendations. *J Consum Res* 2005 Mar;31(4):713-724. [doi: [10.1086/426605](https://doi.org/10.1086/426605)]
68. Schwarz N, Newman E, Leach W. *Behav Sci Policy* 2016;2(1):85-95 [FREE Full text] [doi: [10.1353/bsp.2016.0009](https://doi.org/10.1353/bsp.2016.0009)]
69. Vraga EK, Kim SC, Cook J. Testing logic-based and humor-based corrections for science, health, and political misinformation on social media. *J Broadcast Electron Media* 2019 Sep 20;63(3):393-414. [doi: [10.1080/08838151.2019.1653102](https://doi.org/10.1080/08838151.2019.1653102)]
70. Vraga EK, Bode L. Addressing COVID-19 misinformation on social media preemptively and responsively. *Emerg Infect Dis* 2021 Feb;27(2):396-403. [doi: [10.3201/eid2702.203139](https://doi.org/10.3201/eid2702.203139)] [Medline: [33395379](https://pubmed.ncbi.nlm.nih.gov/33395379/)]

Abbreviations

ANCOVA: analysis of covariance

CDC: Centers for Disease Control and Prevention

CRT: Cognitive Reflection Test

M: misinformation-only experimental condition

MC: misinformation+correction experimental condition

MCR: misinformation+correction+rebuttal experimental condition

MCRC: misinformation+correction+rebuttal+second correction experimental condition

Edited by M Gisondi, J Faust; submitted 09.11.21; peer-reviewed by L Bode, M Waring, S Pesälä; comments to author 27.12.21; revised version received 08.01.22; accepted 13.02.22; published 02.03.22.

Please cite as:

Mourali M, Drake C

The Challenge of Debunking Health Misinformation in Dynamic Social Media Conversations: Online Randomized Study of Public Masking During COVID-19

J Med Internet Res 2022;24(3):e34831

URL: <https://www.jmir.org/2022/3/e34831>

doi: [10.2196/34831](https://doi.org/10.2196/34831)

PMID: [35156933](https://pubmed.ncbi.nlm.nih.gov/35156933/)

©Mehdi Mourali, Carly Drake. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 02.03.2022. 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.

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org

<https://www.jmirpublications.com/>